Complex structure on a real Hilbert space and symplectic structure on a complex Hilbert space

C. S. Sharma
Birkbeck College, University of London, 43 Gordon Square, London WC1H OPD, England

(Received 6 August 1987; accepted for publication 22 December 1987)

Alternative ways of complexifying a real Hilbert space and quaternionizing a complex Hilbert space are described. The work gives some insight into why even though in the finite-dimensional case a complex Hilbert space when viewed as a real Hilbert space and a quaternionic Hilbert space when viewed as a complex Hilbert space have twice their original dimensions, the degrees of freedom of the linear operators remain unchanged. Many ramifications are discussed, among them the reconciliation of the linearity of the adjoint of a semilinear (antilinear) map from one complex Hilbert space to another with the semilinearity (antilinearity) of the adjoint of a semilinear (antilinear) map from one complex Hilbert space to itself. Groundwork is prepared for the study of the noncommutative algebra of additive operators on a quaternionic Hilbert space.

I. INTRODUCTION

Recent works of Horwitz and Biedenharn[1] and Adler[2] have revived interest in the quaternionic Hilbert space as a candidate for modeling certain quantum phenomena. A study of the spectral properties of unitary operators on a quaternionic Hilbert space was undertaken by Coulson and the present author,[3] and the present work owes its origin to some of the questions asked by the referee of Ref. 3. Though those questions were partially answered in that paper, it soon became clear that there exists considerable misunderstanding of the situation and the literature on the subject is far too sparse. A complete and systematic study of the complexification of a real Hilbert space and of the quaternionization of a complex Hilbert space is, therefore, called for, and the present work is an attempt to answer that call.

There are two extreme views on the complexification of a real Hilbert space. It has been claimed by a group of experts, who prefer to remain anonymous, that everything that can be proved about a complex vector space can be proved by regarding it as a real vector space, and "it would make no difference if multiplication of complex numbers had never been invented." On the other hand, a school of thought led by Hawking[4] believes that complexification of manifolds is one of the most fundamental structures discovered by man, and in it lies the key to a real understanding of relativity and phenomena governed by it. It is manifestly clear that although a complex vector space is also real and hence has all the properties which a real vector space has, there are many properties which are peculiar to a complex vector space; furthermore, if a complex vector space is regarded as a real vector space, the dimension, at any rate in the finite-dimensional case, doubles, yet all linear operators retain their original degrees of freedom.

As far as we know, no corresponding views have been expressed about quaternionizing a complex or a real vector space, but peculiarities particular to a quaternionic vector space are likely to be even greater.

There are two ways of complexifying a real vector space. The difference in the two approaches is most readily evident in the finite-dimensional case, where the dimension remains the same in the first approach and halves in the other. There are two corresponding approaches to quaternionizing a complex vector space. We study both these approaches in some detail in the case where the vector space has an inner product defined on it, and in particular we examine what happens to various linear operators defined on the original space.

In the second of the two approaches, the complex structure is defined through a linear isomorphism $\mathbf{i}$ defined on the real vector space in such a way that $\mathbf{i}^2 = -I$, where I is the identity operator. When the vector space is an inner product space $\mathbf{i}$ is an orthogonal operator. In this approach, defining a quaternionic structure, hereafter called a *symplectic structure*, on a complex inner product space requires a corresponding operator $\mathbf{j}$ that is semiunitary (also called antiunitary), that is, an isometry that is semilinear (also called antilinear). Though semilinear operators are increasingly used in quantum theory, we found the definition of the adjoint of such an operator in the footnote of only one[5] of the many books on algebra and quantum theory we consulted. This definition, which requires such an adjoint to be semilinear, seemed to be in stark contradiction with the definition of an adjoint of an additive map, of which both linear and semilinear maps are particular cases, which requires such an adjoint to be linear and was proposed by Pian and this author.[6] We have, therefore, made a deeper study of the problem and resolved the contradiction, which is more apparent than real. We take the view that the definition of Ref. 6 is the more primitive concept from which that of Ref. 5 can be deduced. Noting that an R-module is a linear space in which the scalars are members of a ring R, another reason for looking at the properties of semilinear operators in some detail is that without such operators it is impossible to construct an operator algebra (that is, an R-module that is also a ring) in a quaternionic Hilbert space. The work that follows prepares the groundwork for the study of such an algebra.

For actual applications of quaternionic inner product and Hilbert spaces to quantum theory we refer the reader back to the excellent accounts in Refs. 1 and 2.

 0022-2488/88/051069-10$02.50

II. FORMALITIES

We denote the fields of real and complex numbers by $\mathbb{R}$ and $\mathbb{C}$, respectively, and the skew field of quaternionic numbers by $\mathbb{H}$. Elementary properties of quaternions are described in Ref. 3.

Let $\mathscr{H}$ be a vector space over $\mathbb{F}$ where $\mathbb{F}=\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$. We define a positive definite Hermitian form on $\mathscr{H}$ by

$$\langle\ ,\ \rangle: \mathscr{H}\times\mathscr{H}\to\mathbb{F},$$

$$\langle pu,qv\rangle = p\langle u,v\rangle q^*, \tag{2.1}$$

$$\langle u+v,w\rangle = \langle u,w\rangle + \langle v,w\rangle, \tag{2.2}$$

$$\langle u,v\rangle^* = \langle v,u\rangle, \tag{2.3}$$

$$\langle u,u\rangle = 0 \quad \text{only if} \quad u=0, \tag{2.4}$$

where $p^*=p$ if $\mathbb{F}$ is real, p^* is the *complex conjugate* of p if $\mathbb{F}$ is complex, and p^* is the *quaternionic conjugate* of p if $\mathbb{F}$ is quaternionic.

Let $\mathscr{H}_1$ and $\mathscr{H}_2$ be Hilbert spaces over $\mathbb{F}$. We say that a map $L: \mathscr{H}_1\to\mathscr{H}_2$ is *additive* if and only if, for all $u,v\in\mathscr{H}_1$,

$$L(u+v) = L(u) + L(v). \tag{2.5}$$

If, in addition, the map L satisfies

$$L(pu) = pL(u), \tag{2.6}$$

for all $p\in\mathbb{F}$ and all $u\in\mathscr{H}_1$, then it is called *linear*. If, on the other hand, L satisfies

$$L(pu) = p^*L(u), \tag{2.7}$$

for all $p\in\mathbb{F}$ and all $u\in\mathscr{H}_1$, then it is called *semilinear* or *antilinear*.

It was proved by Pian and this author[7] that in the complex case every additive continuous map from $\mathscr{H}_1$ to $\mathscr{H}_2$ is a direct sum of a linear and a semilinear continuous map from $\mathscr{H}_1$ to $\mathscr{H}_2$. (Reference 7 was in the more general context of Banach spaces of which Hilbert spaces are particular cases.)

A map from a Hilbert space $\mathscr{H}$ to itself will be called an operator on $\mathscr{H}$.

Let A be either a linear or a semilinear operator on $\mathscr{H}$. A norm of A denoted by $\|A\|$ is defined by the formula

$$\|A\| = \sup_{\|x\|=1} \|Ax\|. \tag{2.8}$$

Let $\mathscr{N}_1\oplus\mathscr{N}_2$ be the direct sum of two normed spaces $\mathscr{N}_1$ and $\mathscr{N}_2$. An element of $\mathscr{N}_1\oplus\mathscr{N}_2$ is a pair (x_1,x_2) with $x_1\in\mathscr{N}_1$ and $x_2\in\mathscr{N}_2$. A norm of (x_1,x_2) denoted by $\|(x_1,x_2)\|$ is defined by the formula

$$\|(x_1,x_2)\| = \|x_1\| + \|x_2\|. \tag{2.9}$$

Let A be an additive operator on $\mathscr{H}$. Then by the theorem of Pian and Sharma[7] A belongs to the direct sum of the spaces of linear and semilinear operators on $\mathscr{H}$, and hence by (2.9) its norm is defined by

$$\|A\| = \|A_1\| + \|A_2\|, \tag{2.10}$$

where A_1 and A_2 are, respectively, the linear and semilinear components of A.

III. THE ADJOINT OF A SEMILINEAR OPERATOR ON A COMPLEX HILBERT SPACE

We shall first recall what the Riesz representation theorem tells us. Let us denote the set of all bounded linear functionals on $\mathscr{H}$ by $\tilde{\mathscr{H}}$. This $\tilde{\mathscr{H}}$ is called the dual of $\mathscr{H}$. According to the Riesz representation theorem, elements in $\mathscr{H}$ and $\tilde{\mathscr{H}}$ are related to one another by a norm-preserving semilinear isomorphism. One consequence is that given a bounded linear functional Φ on $\mathscr{H}$, there is a unique vector y in $\mathscr{H}$ such that

$$\Phi = \langle x,y\rangle. \tag{3.1}$$

Given a linear operator A on $\mathscr{H}$ and any fixed vector y in $\mathscr{H}$, Ψ defined by

$$\Psi(x) = \langle Ax,y\rangle \tag{3.2}$$

is clearly a a bounded linear functional on $\mathscr{H}$, and hence to this Ψ there corresponds a unique vector, which we call y^*, such that

$$\langle Ax,y\rangle = \Psi(x) = \langle x,y^*\rangle. \tag{3.3}$$

It is easy to verify that the relation between y and y^* is linear, and we define A^* by

$$A^*y = y^*, \tag{3.4}$$

so that A^* is a linear operator on $\mathscr{H}$. Here A^* is called the adjoint of A. The existence of y^* and therefore of A^* is thus established as a consequence of the Riesz representation theorem.

In more informal works the adjoint A^* of a linear operator A on a Hilbert space $\mathscr{H}$ over $\mathbb{C}$ is defined by

$$\langle A^*u,v\rangle = \langle u,Av\rangle, \tag{3.5}$$

for all $u,v\in\mathscr{H}$. When A is semilinear this definition clearly does not work, because replacing v by αv, where α is a scalar, gives $\langle A^*u,v\rangle\alpha^*$ on the left-hand side but $\langle u,Av\rangle\alpha$ on the right-hand side, and in view of (2.3) the two are equal if and only if α is real.

The adjoint A^* of a semilinear operator A is informally defined by[5]

$$\langle u,A^*v\rangle = \langle v,Au\rangle, \quad \forall u,v\in\mathscr{H}. \tag{3.6}$$

An easy computation shows that A^* is semilinear also. The existence of the adjoint A^* in this case also can be demonstrated by a generalization of the Riesz representation theorem to the semilinear functionals, which we shall presently state and prove. In what follows, the set of bounded semilinear functionals on $\mathscr{H}$ is called the semidual of $\mathscr{H}$ and is denoted by $\tilde{\mathscr{H}}_s$.

Theorem (Riesz representation theorem for bounded semilinear functionals on a Hilbert space): There exists a linear norm preserving isomorphism between a Hilbert space $\mathscr{H}$ and its semidual $\tilde{\mathscr{H}}_s$.

Proof: Let $y\in\mathscr{H}$. Let Φ_y be the semilinear functional on $\mathscr{H}$ defined by

$$\Phi(x) = \langle y,x\rangle. \tag{3.7}$$

We shall show that the correspondence $y\mapsto\Phi_y$ is a norm-preserving linear isomorphism from $\mathscr{H}$ to $\tilde{\mathscr{H}}_s$. The computation

$$\begin{aligned}(y+z)\mapsto\Phi_{y+z}(x) &= \langle y+z,x\rangle = \langle y,x\rangle + \langle z,x\rangle\\ &= \Phi_y(x) + \Phi_z(x)\end{aligned} \tag{3.8}$$

and

$$\alpha y\mapsto\Phi_{\alpha y}(x) = \langle\alpha y,x\rangle = \alpha\langle y,x\rangle = \alpha\Phi(x), \tag{3.9}$$

for every $x \in \mathscr{H}$, shows that the correspondence is linear.

Now

$$\|\Phi_y\| = \sup_{\|x\| = 1} |\langle y,x \rangle| \leq \|y\|, \tag{3.10}$$

but

$$\|\Phi_y\| \geq |\Phi(y/\|y\|)| = |\langle y,y \rangle / \|y\|| = \|y\|. \tag{3.11}$$

Hence

$$\|\Phi_y\| = \|y\| \tag{3.12}$$

and the correspondence is norm-preserving.

Now we must show that if $\Phi \in \tilde{\mathscr{H}}_s$, there exists a unique $y \in \mathscr{H}$ such that $\Phi = \Phi_y$. Let $\mathscr{G} = \ker \Phi$ (that is, the set of vectors mapped to zero by Φ). That Φ is bounded implies that Φ is continuous, which in turn implies that $\mathscr{G}$ is a subspace of $\mathscr{H}$. If $\mathscr{G} = \mathscr{H}$, then $y = 0$ satisfies all the requirements. If $\mathscr{G} \neq \mathscr{H}$, there exists a nonzero vector $z \in \mathscr{G}^{\perp}$. We claim that $y = [\Phi(z)/\|z\|^2]z$ satisfies all our requirements. Remembering that Φ is semilinear, we see that

$$\Phi(x - (\Phi(x)/\Phi(z))^* z) = 0; \tag{3.13}$$

hence $x - (\Phi(x)/\Phi(z))^* z \in \mathscr{G}$ and $z \in \mathscr{G}^{\perp}$, and therefore

$$\langle z, x - (\Phi(x)/\Phi(z))^* z \rangle = 0, \tag{3.14}$$

or

$$\Phi(x) = \langle [\Phi(z)/\|z\|^2] z, x \rangle = \langle y,x \rangle. \tag{3.15}$$

To prove the uniqueness of y, suppose that

$$\Phi(x) = \langle y,x \rangle = \langle y',x \rangle. \tag{3.16}$$

Hence

$$\langle y - y', x \rangle = 0, \quad \forall x \in \mathscr{H}, \tag{3.17}$$

and therefore $y = y'$. This completes the proof.

We now know that given a bounded semilinear functional on $\mathscr{H}$, there exists a unique vector y in $\mathscr{H}$ such that

$$\Phi(x) = \langle y,x \rangle. \tag{3.18}$$

Given a semilinear operator A on $\mathscr{H}$ and any fixed vector y in $\mathscr{H}$, Ψ defined by

$$\Psi(x) = \langle Ax, y \rangle \tag{3.19}$$

is a bounded semilinear functional on $\mathscr{H}$, and hence there exists a unique vector, which we denote by y^*, such that

$$\Psi(x) = \langle y^*, x \rangle. \tag{3.20}$$

An easy verification shows that the relation between y and y^* is semilinear, and we define the adjoint A^* of A by

$$A^* y = y^*. \tag{3.21}$$

Thus given a semilinear operator A, there exists a semilinear operator A^* such that

$$\langle Ax, y \rangle = \langle A^* y, x \rangle, \tag{3.22}$$

which is by complex conjugation equivalent to (3.6).

Now let us consider the adjoint in a more general setup, when $\hat{A}$ is an additive map from one Hilbert space $\mathscr{H}_1$ to another Hilbert space $\mathscr{H}_2$. Let $\mathscr{A}(\mathscr{H},\mathbb{C})$ denote the space of bounded additive maps from $\mathscr{H}$ to $\mathbb{C}$. Letting $f \in \mathscr{A}(\mathscr{H}_2,\mathbb{C})$, then $f \circ \hat{A} \in \mathscr{A}(\mathscr{H}_1,\mathbb{C})$, where $\circ$ denotes the functional composition. The adjoint $\hat{A}^*$ of $\hat{A}$ is a map from $A(\mathscr{H}_2,\mathbb{C})$ to $\mathscr{A}(\mathscr{H}_1,\mathbb{C})$, such that for every $f \in \mathscr{A}(\mathscr{H}_2,\mathbb{C})$,

$$\hat{A}^* f = f \circ \hat{A}. \tag{3.23}$$

In other words, $\hat{A}^*$ is defined in such a way that it makes the following diagram commutative:

$$\begin{tikzcd}
\mathscr{H}_1 \arrow[r, "\hat{A}"] \arrow[d, "\hat{A}^* f"'] & \mathscr{H}_2 \arrow[d, "f"] \\
\mathbb{C} \arrow[r, "\mathrm{id}"'] & \mathbb{C}
\end{tikzcd}$$

Here id is the identity map on $\mathbb{C}$.

Now if

$$f = \alpha_1 f_1 + \alpha_2 f_2, \tag{3.24}$$

then

$$\hat{A}^* f = (\alpha_1 f_1 + \alpha_2 f_2) \circ \hat{A} = \alpha_1 f_1 \circ \hat{A} + \alpha_2 f_2 \circ \hat{A}, \tag{3.25}$$

showing that

$$\hat{A}^*(\alpha_1 f_1 + \alpha_2 f_2) = \alpha_1 \hat{A}^* f_1 + \alpha_2 \hat{A}^* f_2, \tag{3.26}$$

and thus $\hat{A}^*$ is linear.

We shall now show how the definitions (3.5) and (3.6) of adjoints of linear and semilinear operators on a complex Hilbert space can be deduced also from the more general definition (3.23). The Riesz representation theorem and its generalization proved above show that the spaces of bounded linear and semilinear functionals on $\mathscr{H}$ are isomorphic with $\mathscr{H}$, and a simple further generalization shows that the space of bounded additive functionals on $\mathscr{H}$ is isomorphic with $\mathscr{H} \oplus \mathscr{H}$. Thus an additive functional f on $\mathscr{H}$ being the direct sum of a linear and a semilinear functional is represented by two vectors u and v:

$$f(x) = \langle x,u \rangle + \langle v,x \rangle. \tag{3.27}$$

Let $\mathscr{H}_1 = \mathscr{H}_2 = \mathscr{H}$ in the discussion preceding (3.23), and let $\hat{A}$ be linear. Now $\hat{A}$ is an operator on $\mathscr{H}$, and the hat on it has no significance, so we write $A = \hat{A}$. According to our definition, $\hat{A}^*$ is now an operator on the additive dual of $\mathscr{H}$, but according to the Riesz representation theorem the additive dual of $\mathscr{H}$ is isomorphic with $\mathscr{H} \oplus \mathscr{H}$. After transferring $\hat{A}^*$ to $\mathscr{H}$ with the help of this isomorphism we shall drop the hat from $\hat{A}^*$. We shall then find that A^* agrees with its definition in (3.5). In the present setup $f \circ A$ is additive, and its action is given by the formula (3.27), now modified to take into account the action of A:

$$(f \circ A)(x) = \langle Ax,u \rangle + \langle v,Ax \rangle. \tag{3.28}$$

Let f_1 and f_2 be the linear and semilinear components of f. Then from the linearity of $\hat{A}^*$,

$$A^* f = \hat{A}^* f_1 + \hat{A}^* f_2, \tag{3.29}$$

where $\hat{A}^* f_1$, being the composite of two linear maps, is linear, and $\hat{A}^* f_2$, being the composite of a linear and a semilinear map, is semilinear. We now use the Riesz representation theorem to identify f_1, f_2, $\hat{A}^* f_1$, and $\hat{A}^* f_2$ with u, v, $A^* u$, and $A^* v$, respectively. In other words, if the vector u corresponds to the linear functional f_1 under the Riesz representation theorem, we use the symbol $A^* u$ to denote the vector that corresponds to the linear functional $\hat{A}^* f_1$ under the same theorem. It later turns out that this natural symbolism amounts to defining the adjoint of A, which agrees with our earlier definitions. Then we have, with appropriate modification of (3.27),

$$(\hat{A}^*f)(x) = \langle x,A^*u\rangle + \langle A^*v,x\rangle. \tag{3.30}$$

Equating the linear and semilinear components of (3.28) and (3.30) gives definition (3.5) twice over. It is easy to verify that A^* defined in this way is indeed a linear operator on $\mathscr{H}$.

Next suppose that A is a semilinear operator on $\mathscr{H}$. Now

$$\hat{A}^*f_1 = f_1\circ A \tag{3.31}$$

and

$$\hat{A}^*f_2 = f_2\circ A. \tag{3.32}$$

Noting that the composite of a linear and a semilinear map is semilinear while the composite of two semilinear maps is linear, and with the same identifications as before, equations corresponding to (3.28) and (3.30) in the present case take the forms

$$(f\circ A)(x) = \langle v,Ax\rangle + \langle Ax,u\rangle \tag{3.33}$$

and

$$(\hat{A}^*f)(x) = \langle x,A^*v\rangle + \langle A^*u,x\rangle. \tag{3.34}$$

Again equating the linear and semilinear components of the two equations gives us definition (3.6) twice over, and once again it is easy to verify that A^* defined in this way for a semilinear operator is semilinear. Note that when A is linear, $\hat{A}^*$ is a linear operator on the space of bounded additive functionals on $\mathscr{H}$, and takes linear bounded functionals into linear bounded functionals and bounded semilinear functionals to bounded semilinear ones. On the other hand, when A is semilinear, $\hat{A}^*$ is again a linear operator on the space of bounded additive functionals on $\mathscr{H}$, but it now takes bounded linear functionals into bounded semilinear ones and vice versa. Thus we have three definitions of the adjoint maps of linear and semilinear operators, and in each case the last definition is the most primitive and the middle one is least so. According to one definition, the adjoint of a semilinear map is linear and it is semilinear according to the other two, but as we have seen above they are completely consistent with each other though they are different objects being defined on different spaces: on the space of bounded additive functionals on $\mathscr{H}$ that is isomorphic with $\mathscr{H}\oplus\mathscr{H}$ in the last case, and on $\mathscr{H}$ itself in the others. It should be noted that while definitions (3.5) and (3.6) assume the existence of the adjoint with those properties, definitions (3.23), (3.4), and (3.21) establish the existence of these adjoints with the help of the Riesz representation theorem and its generalization to the space of bounded semilinear functionals on $\mathscr{H}$.

It will be shown in a later section that the concept of additivity takes on a very important role when we are dealing with quaternionic Hilbert spaces. In these spaces neither linear functionals nor linear operators form a linear space. The Riesz representation theorem as a vector space isomorphism between a vector space and its dual has no meaning in quaternionic Hilbert spaces. However, it was shown by Horwitz and Biedenharn[1] that the Riesz representation theorem in a more restricted sense is still valid in quaternionic Hilbert spaces, and the restricted theorem is adequate to guarantee the existence of the adjoint of any linear operator on such a space. In quaternionic Hilbert spaces, for every linear operator there are three semilinear ones, and they are in fact obtained from each other by multiplying the operator by a scalar—a fact that enables one to extend the definition of the adjoint to semilinear operators and finally to all additive operators. Additive operators on a quaternionic Hilbert space are the only family of operators that form a linear space, an R-module (a linear space in which scalars are members of a ring rather than a field) to be more precise, over the quaternions. They also form a ring and thus constitute a noncommutative algebra. Before quaternionic Hilbert spaces can play their due role in physics, the algebra of additive operators must be studied and understood. A study of semilinear and additive operators on a complex Hilbert space is, in the opinion of the present author, a necessary step in this direction. The present study has already helped us in starting the development of a systematic study of bounded additive operators on a complex Hilbert space.[8]

IV. COMPLEXIFICATION OF A REAL VECTOR SPACE WITHOUT LOSS OF DIMENSION

This construction is more fully described in Halmos.[9] We recapitulate here the main features so that we can generalize them to the problem of quaternionizing a complex vector space. There are three equivalent but different ways of formulating this version of complexification.

(i) The easiest, though naive, way of doing this is to enlarge the underlying set of a vector $\mathscr{V}$ to a set $\mathscr{V}_c$. This set $\mathscr{V}_c$ includes each element v of $\mathscr{V}$. In addition, for each element v in $\mathscr{V}$, an extra element called iv is supposed to exist in $\mathscr{V}_c$. The vector iv is supposed to be linearly dependent on v in $\mathscr{V}_c$ regarded as a complex space, and is postulated to have the property

$$i\cdot iv = -v. \tag{4.1}$$

This gives us the really pedestrian way of complexifying a real vector space. The same effect is achieved by two more rigorous but equivalent formulations.

(ii) The complexification $\mathscr{V}_c$ of a real vector space $\mathscr{V}$ is defined to be a vector space structure on the set $\mathscr{V}\times\mathscr{V}$ over the field $\mathbb{R}\times\mathbb{R}$ identified as $\mathbb{C}$, with vector addition and scalar multiplication defined, for $u_1,u_2,v_1,v_2\in\mathscr{V}$, by

$$(u_1,u_2) + (v_1,v_2) = (u_1+v_1,u_2+v_2), \tag{4.2}$$

and, for $r_1,r_2\in\mathbb{R}$ and $u_1,u_2\in\mathscr{V}$, by

$$(r_1,r_2)\cdot(u_1,u_2) = (r_1\cdot u_1 - r_2\cdot u_2, r_1\cdot u_2 + r_2\cdot u_1). \tag{4.3}$$

(iii) The complexification $\mathscr{V}_c$ of a real vector space $\mathscr{V}$ is defined to be the tensor product space $\mathbb{C}\otimes\mathscr{V}$, where $\mathbb{C}$ is regarded as a two-dimensional real vector space. Letting $\beta\otimes v$ be a typical element of $\mathbb{C}\otimes\mathscr{V}$, we define scalar multiplication of $\beta\otimes v$ by $\alpha\in\mathbb{C}$ by

$$\alpha\cdot(\beta\otimes v) = (\alpha\cdot\beta)\otimes v, \tag{4.4}$$

where $\alpha\cdot\beta$ is that element in $\mathbb{C}$ that corresponds to the product of α and β as complex numbers.

Superficially the three constructions look wildly different, but it is relatively easy to prove that they are all equivalent. The main features of these constructions are as follows.

(a) The dimension of the complexified space $\mathscr{V}_c$ as a complex space is the same as that of the original space $\mathscr{V}$ as a real space.

(b) There is a natural way of extending any linear operator A on $\mathscr{V}$ to a linear operator A_c on $\mathscr{V}_c$:

$$A_c(u + iv) = Au + iAv. \tag{4.5}$$

A similar extension works for linear and even multilinear functionals. In particular, the inner product, which is a real bilinear functional on $\mathscr{V}$, is naturally extended to a complex (Hermitian) inner product by

$$\langle u_1 + iv_1, u_2 + iv_2\rangle_c = \langle u_1,u_2\rangle + \langle v_1,v_2\rangle + i(-\langle u_1,v_2\rangle + \langle u_2,v_1\rangle). \tag{4.6}$$

Thus the Hilbert space structure easily goes over into the complexified space.

(c) The correspondence $A \mapsto A_c$ preserves all algebraic properties of the operator. Here are examples.

(1) If $B = rA$ with r real, then $B_c = rA_c$.

(2) If $C = A + B$, then $C_c = A_c + B_c$.

(3) If $C = AB$, then $C_c = A_c B_c$.

(4) If $\mathscr{V}$ is an inner product space and $B = A^*$, then $B_c = A_c^*$.

(5) If the complexification A_c of a linear operator A has an eigenvector $u + iv$ with eigenvalue $r + is$ ($r,s \in \mathbb{R}$), then

$$\begin{aligned} A_c(u + iv) &= (r + is)(u + iv) \\ &= ru - sv + i(su + rv). \end{aligned} \tag{4.7}$$

Hence, from the definition (4.5) of A_c,

$$Au = ru - sv, \tag{4.8}$$

$$Av = su + rv. \tag{4.9}$$

Thus the subspace of $\mathscr{V}$ spanned by u and v is invariant under A. If the eigenvalue is real, that is, if $s = 0$, then $Au = ru$ and $Av = rv$, and since by definition an eigenvector is nonzero, both u and v cannot vanish. Since every linear operator on a complex space has at least one eigenvector, we conclude in passing that every linear operator on a real vector space leaves invariant a subspace of dimension 1 or 2.

(d) Every basis in $\mathscr{V}$ is a basis in $\mathscr{V}_c$, and both A and A_c with respect to a particular basis of this kind are represented by the same real matrix.

(e) The dimension of the complexified space $\mathscr{V}_c$ regarded as a real space $\mathscr{V}_{cr}$ is, in the finite-dimensional case, twice the dimension of the original space, and in all cases the original space $\mathscr{V}$ is isomorphic with a proper subspace of $\mathscr{V}_{cr}$, which is the complexified space $\mathscr{V}_c$ regarded as a real space.

We shall discuss the relationship between $\mathscr{V}_c$ and $\mathscr{V}_{cr}$ in somewhat greater detail because here lies the basis of the second kind of complexification.

Once the complexified space $\mathscr{V}_c$ is regarded as a real space $\mathscr{V}_{cr}$, though objects denoted by iu ($u \in \mathscr{V}$) are members of $\mathscr{V}_{cr}$, i as the square root of -1 has no place in $\mathscr{V}_{cr}$. Fortunately, the relation between u and iu is a linear one, that is, if $u = rv + sw$ ($r,s \in \mathbb{R}$), then

$$iu = riv + siw, \tag{4.10}$$

and we can and shall regard i as a linear operator on $\mathscr{V}_{cr}$, and denote it by a bold $\mathbf{i}$. It is easy to see that the linear operator $\mathbf{i}$ has the property

$$\mathbf{i}^2 = -I, \tag{4.11}$$

where I is the identity map on $\mathscr{V}_{cr}$.

Earlier in this section we saw that every linear map A on $\mathscr{V}$ has a natural extension A_c as a linear map on $\mathscr{V}_c$. When $\mathscr{V}_c$ is regarded as a real space $\mathscr{V}_{cr}$, each of the extended maps A_c continues to be linear, and we shall denote it by A_{cr} when it is regarded as a linear map on $\mathscr{V}_{cr}$. Each such linear operator commutes with the linear operator $\mathbf{i}$:

$$A_{cr}\mathbf{i} = \mathbf{i}A_{cr}. \tag{4.12}$$

One immediate consequence is that if u is an eigenvector of A_{cr}, then so is $\mathbf{i}u$, and thus every point in the spectrum of A is at least doubly degenerate. Given any basis $\mathscr{B}$ in $\mathscr{V}_c$, Span $\mathscr{B}$ in $\mathscr{V}_{cr}$ is a proper subspace of $\mathscr{V}_{cr}$, and $\mathscr{V}_{cr}$ has the decomposition

$$\mathscr{V}_{cr} = \text{Span } \mathscr{B} \oplus \mathbf{i}(\text{Span } \mathscr{B}). \tag{4.13}$$

The linear operator $\mathbf{i}$ being a linear bijection is a vector space isomorphism from $\mathscr{V}_{cr}$ to itself, and, for any basis $\mathscr{B}$ in $\mathscr{V}_c$, restriction of $\mathbf{i}$ to Span $\mathscr{B}$ in $\mathscr{V}_{cr}$ is an isomorphism from Span $\mathscr{B}$ to $\mathbf{i}$(Span $\mathscr{B}$).

We have seen that if there is an inner product defined on $\mathscr{V}$, it has a natural extension as a Hermitian inner product on $\mathscr{V}_c$. This Hermitian product induces a real inner product on $\mathscr{V}_{cr}$:

$$\langle u,v\rangle_{cr} = \text{real part of } \langle u,v\rangle_c. \tag{4.14}$$

An immediate consequence of this is that for any vector u, $\mathbf{i}u$ is perpendicular to u.

If $\mathscr{V}_c$ has dimension 1 (that is, $\mathscr{V}_c$ is $\mathbb{C}$ regarded as a vector space complexification of $\mathbb{R}$) with the usual norm, then $\mathbf{i}$ in $\mathscr{V}_{cr}$, which is two-dimensional, has the matrix representation

$$\mathbf{i} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{4.15}$$

It is again easy to verify that $\mathbf{i}$ is orthogonal and satisfies Eq. (4.11). For the general case, a matrix representation for $\mathbf{i}$ can be obtained by placing the matrix in (4.15) as block matrices along the principal diagonal and placing zeros everywhere else, but this representation is not unique.

Even though the dimension of $\mathscr{V}_{cr}$ is, in the finite-dimensional case, twice the dimension of $\mathscr{V}_c$ or $\mathscr{V}$—because of the isomorphism between span $\mathscr{B}$ in $\mathscr{V}_{cr}$, where $\mathscr{B}$ is any basis in $\mathscr{V}_c$, and one of its complements referred to above, and because every linear operator A_{cr} on $\mathscr{V}_{cr}$, which remains linear when the same space is regarded as a complex space $\mathscr{V}_c$, commutes with $\mathbf{i}$—it is enough to know A_{cr} on a proper subspace (or half the dimension of $\mathscr{V}_{cr}$ in the finite-dimensional case) to determine its behavior on the entire space. This is what we mean when we say that the degree of freedom of a linear operator does not increase when we regard $\mathscr{V}_c$ as a real space $\mathscr{V}_{cr}$. This also explains why, in the functional calculus of variations on a complex Hilbert space, equating to zero the partial derivatives of a functional with respect to real and complex parts of a complex parameter treated as two independent variables does not, in general, yield two independent equations (see Ref. 6 for further discussion).

Every linear functional f on the original real space $\mathscr{V}$ has a natural extension to its complexification $\mathscr{V}_c$ by

$$f_c(iu) = if(u). \tag{4.16}$$

Furthermore, if can itself be regarded as a linear functional on $\mathscr{V}_c$, with f and if belonging to the same one-dimensional subspace in the dual of $\mathscr{V}_c$. In $\mathscr{V}_{cr}$, f_c induces a linear functional f_{cr} on $\mathscr{V}_{cr}$ by

$$f_{cr}(u) = \text{real part of } f_c(u), \tag{4.17}$$

and in $\mathscr{V}_{cr}$, f_{cr} and $(if)_{cr}$ are no longer linearly dependent.

V. QUATERNIONIZING A COMPLEX VECTOR SPACE

Just as any complex number can be written as a pair of real numbers with sums and products defined by

$$(r_1,s_1) + (r_2,s_2) = (r_1 + r_2, s_1 + s_2), \tag{5.1}$$

$$(r_1,s_1)\cdot(r_2,s_2) = (r_1r_2 - s_1s_2, r_1s_2 + r_2s_1), \tag{5.2}$$

any quaternionic number can be written as a pair of complex numbers with sums and products defined by

$$(c_1d_1) + (c_2d_2) = (c_1 + c_2, d_1 + d_2), \tag{5.3}$$

$$(c_1,d_1)\cdot(c_2,d_2) = (c_1c_2 - d_1{}^*d_2, c_1{}^*d_2 + c_2d_1). \tag{5.4}$$

Keeping in mind the small differences in these definitions, it is easy to see that all three methods described in the previous section for complexifying a real vector space have obvious generalizations for quaternionizing a complex vector space $\mathscr{V}$ (note that in this section $\mathscr{V}$ without a subscript is a complex space).

(i) For each vector v in $\mathscr{V}$ we add a vector jv to the underlying set and regard the enlarged set as a quaternionic space, with jv playing the role of the vector resulting from v as a result of scalar multiplication by j.

(ii) We take $\mathscr{V}\times\mathscr{V}$ as the underlying set of a vector space over $\mathbb{C}\times\mathbb{C}$, which is identified with $\mathbb{H}$, with vector sum and scalar multiplication defined by

$$(u_1,u_2) + (v_1,v_2) = (u_1 + v_1, u_2 + v_2), \tag{5.5}$$

$$(c_1,c_2)\cdot(u_1,u_2) = (c_1u_1 - c_2{}^*u_2, c_1{}^*u_2 + c_2u_1). \tag{5.6}$$

(iii) We regard $\mathbb{H}$ as a two-dimensional vector space over $\mathbb{C}$, and regard the tensor product $\mathbb{H}\otimes\mathscr{V}$ as a quaternionic vector space with scalar product defined by

$$p\cdot(q\otimes u) = (p\cdot q)\otimes u, \tag{5.7}$$

where $q\otimes u$ is a typical element of $\mathbb{H}\otimes\mathscr{V}$ and $(p\cdot q)$ is the product of p and q as quaternionic numbers.

As in Sec. IV, it is easy to show that the three constructions are completely equivalent. In each case we denote the original space by $\mathscr{V}$ and its quaternionization by $\mathscr{V}_q$. Remember that we now have three anticommuting, linearly independent square roots of -1, namely i, j, and k, satisfying

$$ij = -ji = k \tag{5.8}$$

and that in each of the above constructions k is written in terms of i and j in accordance with (5.8). Much of what has been described in the preceding section carries through for the construction of this section. We shall, therefore, concentrate our attention on points where the situation is somewhat different. The differences arise because of the lack of commutativity between i and j. When $\mathscr{V}$ has an inner product as a complex space, the inner product is extended to $\mathscr{V}_q$ by the formula

$$\langle u_1 + jv_1, u_2 + jv_2\rangle_q = \langle u_1,u_2\rangle + \langle v_2,v_1\rangle + j(\langle v_1,u_2\rangle - \langle v_2,u_1\rangle). \tag{5.9}$$

As was pointed out in Ref. 3, another consequence of the lack of commutativity in $\mathbb{H}$ is that scalar multiples of linear operators (or functionals) are not linear operators (or functionals) in $\mathscr{V}_q$. Thus neither the linear operators nor the linear functionals on $\mathscr{V}_q$ form a vector space (or R-module), and the Riesz representation theorem which asserts that every complex Hilbert space is isomorphic with its dual becomes meaningless if stated in those terms for a Hilbert space over $\mathbb{H}$. However, it was shown by Horwitz and Biedenharn[1] that a more restricted version of the Riesz representation theorem still holds for Hilbert spaces over $\mathbb{H}$ and, fortunately, that is enough to prove the existence of the adjoint of a linear operator as a linear operator on the space.

When $\mathscr{V}_q$ is regarded as a complex space $\mathscr{V}_{qc}$, the original space $\mathscr{V}$ (which in this section is complex) is isomorphic with a proper subspace of $\mathscr{V}_{qc}$, and in the finite-dimensional case the dimension of $\mathscr{V}_{qc}$ is twice that of $\mathscr{V}$, as in the previous section. In $\mathscr{V}_{qc}$, u and ju are linearly independent vectors, but in this space j is neither a vector nor a scalar but, in analogy with the situation in the preceding section, can be regarded as an operator, though because of lack of commutativity between i and j, it is no longer linear. It is easy to verify that it is a semilinear (antilinear) automorphism on $\mathscr{V}_{qc}$ with the property

$$-\mathbf{j}^2 = I, \tag{5.10}$$

where, as in the preceding section, we have denoted j as an operator by a bold letter, and I is the identity map on $\mathscr{V}_{qc}$. Because complex matrices represent linear operators on a complex space, $\mathbf{j}$ does not have a matrix representation.

It has already been pointed out that by writing k in terms of i and j, every quaternionic number q can be written in the form

$$q = c_1 + jc_2, \tag{5.11}$$

where c_1 and c_2 are complex numbers of the form

$$c_n = a_n + ib_n \quad (n = 1,2). \tag{5.12}$$

However, it must be pointed out that the choice of neither i nor j is unique, but once i and j are chosen, k becomes unique with respect to this choice. Once this choice has been made (there are infinitely many possible choices), we can call c_1 the complex part of q. Now in going from $\mathscr{V}_q$ to $\mathscr{V}_{qc}$, we can define an inner product on $\mathscr{V}_{qc}$, if an inner product already exists on $\mathscr{V}_q$, by

$$\langle u,v\rangle_{qc} = \text{complex part of } \langle u,v\rangle_q. \tag{5.13}$$

It seems to the present author that the difference that is most significant and probably most pregnant with possibilities for further develoment is the following. In the previous section, $\mathscr{V}$ was a proper subspace of $\mathscr{V}_{cr}$; for each vector u in $\mathscr{V}$ there was an extra vector iu in $\mathscr{V}_{cr}$ linearly independent of u; and similarly, for each linear functional f on $\mathscr{V}$ there was an extra linear functional if on $\mathscr{V}_{cr}$. In the present case, though, for each u in $\mathscr{V}$ there is an extra vector ju in $\mathscr{V}_{qc}$, and for each linear functional f on $\mathscr{V}$, jf is not even linear on $\mathscr{V}_{qc}$. Yet according to the Riesz representation theorem (isomorphism between $\mathscr{V}_{qc}$ and its dual) there must be extra linear functionals on $\mathscr{V}_{qc}$. What are these? The answer, though quite simple, is rich in possibilities. As has been stated earlier, the space of semilinear functionals is also isomorphic with the original space; for each semilinear functional g on $\mathscr{V}$, $\mathbf{j}g$ is a linear functional on $\mathscr{V}_{qc}$. It was shown in Ref. 7 that the space of bounded additive functionals on a Banach space is a direct sum of the spaces of bounded linear and semilinear functionals on that space. Thus there is a certain isomorphism between the space $\mathscr{A}(\mathscr{V},\mathbb{C})$ of bounded additive functionals on $\mathscr{V}$ and the space $\mathscr{L}(\mathscr{V}_{qc},\mathbb{C})$ of bounded linear functionals on $\mathscr{V}_{qc}$. Here again as in Sec. III we have a correspondence, through an isomorphism, between a semilinear functional on $\mathscr{V}$ and a linear one on $\mathscr{V}_{qc}$.

As an example of application of the interplay between $\mathscr{V}_{qc}$ and $\mathscr{V}_q$, we shall demonstrate the existence of the adjoint of a linear map on $\mathscr{V}_q$ by taking it into $\mathscr{V}_{qc}$. We know that each linear map on $\mathscr{V}_q$ induces a linear map on $\mathscr{V}_{qc}$ that commutes with $\mathbf{j}$, and only those linear maps on $\mathscr{V}_{qc}$ that commute with $\mathbf{j}$ remain linear on $\mathscr{V}_q$. Each linear operator A on $\mathscr{V}_q$ is a linear operator on $\mathscr{V}_{qc}$ that commutes with $\mathbf{j}$. Since $\mathscr{V}_{qc}$ is a complex space we know that its adjoint A^* exists and is a linear operator. The following calculation shows that A^* commutes with $\mathbf{j}$ and therefore is a linear operator on $\mathscr{V}_q$ also:

$$\begin{aligned}\langle u,A^*\mathbf{j}v\rangle &= \langle Au,\mathbf{j}v\rangle = \langle v,\mathbf{j}^*Au\rangle = -\langle v,\mathbf{j}Au\rangle \\ &= -\langle v,A\mathbf{j}u\rangle = -\langle A^*v,\mathbf{j}u\rangle \\ &= -\langle u,\mathbf{j}^*A^*v\rangle = \langle u,\mathbf{j}A^*v\rangle,\end{aligned} \tag{5.14}$$

where we have used definition (3.5) for the adjoint of the linear operator A, definition (3.6) for the adjoint of the semilinear operator $\mathbf{j}$, and the fact that $\mathbf{j}$ is a semilinear automorphism on $\mathscr{V}_{qc}$, that is,

$$\mathbf{j}^* = -\mathbf{j}. \tag{5.15}$$

VI. QUATERNIONIZING A REAL VECTOR SPACE

The methods of the preceding sections can be used to quaternionize directly a real vector space. A quaternionic number can be regarded as a four-dimensional vector space over $\mathbf{R}$ on which a multiplication is defined by

$$\begin{aligned}&(a_0,a_1,a_2,a_3)\cdot(b_0,b_1,b_2,b_3)\\ &\quad= (a_0b_0 - a_1b_1 - a_2b_2 - a_3b_3,\\ &\qquad a_0b_1 + a_1b_0 + a_2b_3 - a_3b_2,\\ &\qquad a_0b_2 + a_2b_0 + a_3b_1 - a_1b_3,\\ &\qquad a_0b_3 + a_3b_0 + a_1b_2 - a_2b_1).\end{aligned} \tag{6.1}$$

All the three equivalent methods described in the preceding sections produce a quaternionic vector space $\mathscr{V}_q$ from a real vector space $\mathscr{V}$.

(i) For each vector u in $\mathscr{V}$ we add vectors iu, ju, and ku to the underlying set. These vectors play the role of scalar multiples of u by i, j, and k, respectively, and we regard the enlarged set as a vector space over $\mathbf{H}$.

(ii) We regard $\mathscr{V}\times\mathscr{V}\times\mathscr{V}\times\mathscr{V}$ as the underlying set of a vector space over $\mathbf{R}^4$ (identified as $\mathbf{H}$) with scalar multiplication defined by

$$\begin{aligned}&(a_0,a_1,a_2,a_3)\cdot(u_0,u_1,u_2,u_3)\\ &\quad= (a_0u_0 - a_1u_1 - a_2u_2 - a_3u_3,\\ &\qquad a_0u_1 + a_1u_0 + a_2u_3 - a_3u_2,\\ &\qquad a_0u_2 + a_2u_0 + a_3u_1 - a_1u_3,\\ &\qquad a_0u_3 + a_3u_0 + a_1u_2 - a_2u_1).\end{aligned} \tag{6.2}$$

(iii) We regard $\mathbf{H}$ as a four-dimensional vector space over $\mathbf{R}$ and regard the tensor product $\mathbf{H}\otimes\mathscr{V}$ as the quaternionic space $\mathscr{V}_q$ with scalar multiplication defined by

$$p\cdot(q\otimes u) = (p\cdot q)\otimes u, \tag{6.3}$$

where $q\otimes u$ is a typical element of $\mathscr{H}\otimes\mathscr{V}$ and $p\cdot q$ is the product of p and q regarded as quaternionic numbers.

It is easy to see that all the three methods are equivalent. Properties of the construction are similar to those in Sec. IV. We will, therefore, mention only the features or formulas where there is a significant difference. When $\mathscr{V}$ has an inner product, the inner product can be extended to $\mathscr{V}_q$ by the following formula, with $u_0,u_1,u_2,u_3,v_0,v_1,v_2,v_3\in\mathscr{V}$:

$$\begin{aligned}&\langle u_0 + iu_1 + ju_2 + ku_3, v_0 + iv_1 + jv_2 + kv_3\rangle_q\\ &\quad= \langle u_0,v_0\rangle + \langle u_1,v_1\rangle + \langle u_2,v_2\rangle + \langle u_3,v_3\rangle\\ &\qquad + i(\langle u_1,v_0\rangle - \langle u_0,v_1\rangle - \langle u_2,v_3\rangle + \langle u_3,v_2\rangle)\\ &\qquad + j(\langle u_2,v_0\rangle - \langle u_0,v_2\rangle - \langle u_3,v_1\rangle + \langle u_1,v_3\rangle)\\ &\qquad + k(\langle u_3,v_0\rangle - \langle u_0,v_3\rangle - \langle u_1,v_2\rangle + \langle u_2,v_1\rangle).\end{aligned} \tag{6.4}$$

For each linear functional f on $\mathscr{V}$, the functionals if, jf, and kf on $\mathscr{V}$ are not linear (cf. Sec. V). So $\mathscr{V}_q$ can be regarded as a real space $\mathscr{V}_{qr}$, and if there is an inner product on $\mathscr{V}_q$, an inner product on $\mathscr{V}_{qr}$ is defined by simply taking the real part of the inner product on $\mathscr{V}_q$. In $\mathscr{V}_{qr}$, i, j, and k are orthogonal operators, and as in earlier sections they will, in this role, be denoted by bold letters. Further, they have the property

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -I, \tag{6.5}$$

where I is the identity operator on $\mathscr{V}_{qr}$. In the simple case when $\mathscr{V}_q$ is one-dimensional over $\mathbf{H}$, a matrix representation for $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ is given by

$$\mathbf{i} = \begin{bmatrix} 0 & -1 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & -1 & 0\end{bmatrix}, \tag{6.6}$$

$$\mathbf{j} = \begin{bmatrix} 0 & 0 & -1 & 0\\ 0 & 0 & 0 & -1\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\end{bmatrix}, \tag{6.7}$$

$$\mathbf{k} = \begin{bmatrix} 0 & 0 & 0 & 1\\ 0 & 0 & -1 & 0\\ 0 & 1 & 0 & 0\\ -1 & 0 & 0 & 0\end{bmatrix}. \tag{6.8}$$

This matrix representation is, however, not unique. In the general case a matrix representation for each of these operators can be constructed by placing the above representation as block matrices along the principal diagonal and zeros everywhere else. Again the representation is not unique.

VII. COMPLEXIFICATION OF A REAL VECTOR SPACE WITH LOSS OF DIMENSION

This is the complexification described by Wells[10] and Chern,[11] and it is the converse of the process by which $\mathscr{V}_{cr}$ is obtained from $\mathscr{V}_c$ (cf. Sec. IV). According to these authors, a *complex structure* on a real vector space $\mathscr{V}$ is an isomorphism $\mathbf{i}$ with the property that

$$\mathbf{i}^2 = -I, \tag{7.1}$$

where I is the identity map on $\mathscr{V}$. The real vector space $\mathscr{V}$ becomes a complex vector space $\mathscr{V}_c$ as soon as $\mathbf{i}u$, for any $u \in \mathscr{V}$, is identified with $i \cdot u$ where $i = \sqrt{-1}$. When a complex structure exists, it is possible to decompose $\mathscr{V}$ into a direct sum of subspaces $\mathscr{U}$ and $\mathscr{W}$,

$$\mathscr{V} = \mathscr{U} \oplus \mathscr{W}, \tag{7.2}$$

such that

$$\mathbf{i}(\mathscr{U}) = \mathscr{W} \tag{7.3}$$

and

$$\mathbf{i}(\mathscr{W}) = \mathscr{U}. \tag{7.4}$$

Because of this decomposition, if a linear operator A commutes with $\mathbf{i}$, it is enough to know the restriction of A on $\mathscr{U}$ or $\mathscr{W}$ to know all its properties. Thus linear operators consistent with the complex structure have only half the degree of freedom compared to an arbitrary linear operator on $\mathscr{V}$. For inner product spaces, the decomposition (7.2) is in orthogonal subspaces while $\mathbf{i}$ is an isometry and, therefore, orthogonal (note that a unitary transformation—a transformation whose adjoint is equal to its inverse—on a real Hilbert space is called orthogonal). The operator $\mathbf{i}$ is not unique, but once it has been chosen, the inner product on $\mathscr{V}$ induces an inner product on $\mathscr{V}_c$,

$$\langle u,v\rangle_c = -\langle u,v\rangle - i\langle \mathbf{i}u,\mathbf{v}\rangle, \tag{7.5}$$

where u,v are any two vectors in $\mathscr{V}$. Furthermore, for a normal operator A that commutes with $\mathbf{i}$, there exists a decomposition of $\mathscr{V}$ into mutually orthogonal subspaces $\mathscr{U}$ and $\mathscr{W}$ such that both $\mathscr{U}$ and $\mathscr{W}$ reduce A. Thus for this most important class of normal operators consistent with the complex structure, each operator is effectively defined on a proper subspace (of half the dimension in the finite-dimensional case). Another consequence is that each point in the spectrum of such an operator has to be at least doubly degenerate.

VIII. QUATERNIONIZING A COMPLEX VECTOR SPACE

This was briefly described by Coulson and the present author in Ref. 3. Things are very much the same except that in place of the linear operator $\mathbf{i}$ we now have a semilinear operator $\mathbf{j}$, though properties (7.1)–(7.4) still hold with $\mathbf{i}$ replaced by $\mathbf{j}$. If an inner product exists on the original complex space $\mathscr{V}$, then once $\mathbf{j}$ has been chosen an inner product is induced on the quaternionized space $\mathscr{V}_q$ by the formula

$$\langle u,v\rangle_q = \langle u,v\rangle - j\langle \mathbf{j}u,\mathbf{v}\rangle, \tag{8.1}$$

where u and v are any two vectors in $\mathscr{V}$. For inner product spaces $\mathbf{j}$ is a semilinear isometry and, therefore, semiunitary. Again all linear and (semilinear) operators on $\mathscr{V}$ that continue to be linear (and semilinear) on $\mathscr{V}_q$ must commute with $\mathbf{j}$, and it is enough to know them on a proper subspace (of half the dimension in the finite-dimensional case) to know them completely. Once again, normal operators consistent with this *symplectic structure* must have each point in their spectra at least doubly degenerate. As has been stated before, the operator $\mathbf{j}$ being semilinear does not have a simple matrix representation.

IX. QUATERNIONIZING A REAL VECTOR SPACE WITH LOSS OF DIMENSION

In this case we need three isomorphisms $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ with the properties

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -I \tag{9.1}$$

and

$$\mathbf{ij} = -\mathbf{ji} = \mathbf{k}. \tag{9.2}$$

By identifying $\mathbf{i}u$, $\mathbf{j}u$, and $\mathbf{k}u$ with iu, ju, and ku, respectively, where i, j, and k are the quaternionic square roots of -1, a real vector space $\mathscr{V}$ is turned into a quaternionic vector space $\mathscr{V}_q$. In this case it is possible to find a subspace $\mathscr{U}$ such that $\mathscr{V}$ has the decomposition

$$\mathscr{V} = \mathscr{U} \oplus \mathbf{i}(\mathscr{U}) \oplus \mathbf{j}(\mathscr{U}) \oplus \mathbf{k}(\mathscr{U}). \tag{9.3}$$

Only those linear operators that commute with the operators $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ remain linear on $\mathscr{V}_q$. For them it is enough to know their behavior on $\mathscr{U}$ to know their behavior on the entire space, so they have effectively a degree of freedom that, in the finite-dimensional case, is a fourth of the dimension of $\mathscr{V}$. In the case where $\mathscr{V}$ has an inner product, $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are orthogonal; a possible matrix representation for them was indicated at the end of Sec. VI. For a particular choice of these operators (the choice is not unique) an inner product is induced on $\mathscr{V}_q$ by the inner product of $\mathscr{V}$:

$$\langle u,v\rangle_q = \langle u,v\rangle - \sum_\tau \tau\langle \boldsymbol{\tau}u,v\rangle, \tag{9.4}$$

where u and v are any two vectors in $\mathscr{V}$ and $\tau = i, j$, or k.

X. THE ALGEBRA OF OPERATORS

Though the term *algebra* is often used to mean other structures, it is common practice among algebraists at the present time to mean by the term *algebra* an R-module (a linear space in which the scalars are members of a ring R) that is also a ring. If the ring R over which such a module is defined is commutative, then the algebra is said to be commutative.

In the case of a Hilbert space over $\mathbb{R}$, the bounded linear operators constitute a normed $*$-algebra ($*$ is the operation by which adjoints are formed and is an involution).

In the case of a Hilbert space over $\mathbb{C}$, it is easy to verify that the bounded linear operators constitute a normed $*$-algebra that is a subalgebra of a bigger normed $*$-algebra of bounded additive operators. The space of bounded additive operators is the direct sum of the spaces of bounded linear and bounded semilinear operators. The bounded semilinear operators, unlike the linear ones, do not form a subalgebra. Furthermore, for two semilinear operators A and B and two complex numbers c and d we have

$$(cA)(dB) = cd^*AB. \tag{10.1}$$

Because of this peculiar property, the algebra of bounded additive operators is a particular kind of commutative algebra. As far as we know, little work has been done on the study of this algebra. We believe that it is important both for applications in quantum mechanics and for developing the algebra of operators on a quaternionic Hilbert space, and we are now engaged in a detailed and systematic study of this algebra.

In the case of a Hilbert space over $\mathbb{H}$, bounded linear operators do not form even a module, let alone an algebra. However, there exists a normed $*$-algebra of bounded additive operators on such a Hilbert space and linear operators are members of this algebra, but unlike the complex case they do not form a subalgebra.

It was proved by Pian and the present author[7] that every bounded additive operator on a Hilbert space over $\mathbb{C}$ is a sum of a bounded linear and a bounded semilinear operator. There exists a similar decomposition of a bounded additive operator on a Hilbert space over $\mathbb{H}$, but before we can state it we need a definition.

Definition: An additive operator A on a Hilbert space $\mathscr{H}$ over $\mathbb{H}$ is said to be *i-semilinear* if and only if it has the following properties: for every r in $\mathbb{R}$ and for every u in $\mathscr{H}$,

$$A(ru) = rAu, \tag{10.2}$$

$$A(iu) = iAu, \tag{10.3}$$

$$A(ju) = -jAu, \tag{10.4}$$

$$A(ku) = -kAu, \tag{10.5}$$

where i, j, and k are as in Sec. II.

Compared to the complex case, the semilinearity is with respect to j and k, yet we call this i-semilinear. This may seem odd, but if the complex part of a quaternionic number is defined with respect to i, then the pure quaternionic part is defined to be j (or k) times a complex number, and the definition of i-semilinearity,

$$A(c+jd)u = cAu - jdAu, \tag{10.6}$$

parallels the definition of semilinearity in the case of operators on a complex Hilbert space. Thus i in i-semilinear indicates that in the decomposition of a quaternionic number q as

$$q = c + \tau d, \tag{10.7}$$

c depends on i alone and τ is either j or k. There are analogous definitions of j-and k-semilinearities. These are obtained by cyclic permutations of i, j, and k in the definition of i-semilinearity.

It was shown by Coulson[12] that a bounded additive operator A on a quaternionic Hilbert space $\mathscr{H}$ can be written as a sum of four operators:

$$A = A_0 + A_1 + A_2 + A_3, \tag{10.8}$$

where

$$A_0u = \tfrac{1}{4}[Au - iA(iu) - jA(ju) - kA(ku)], \tag{10.9a}$$

$$A_1u = \tfrac{1}{4}[Au - iA(iu) + jA(ju) + kA(ku)], \tag{10.9b}$$

$$A_2u = \tfrac{1}{4}[Au + iA(iu) - jA(ju) + kA(ku)], \tag{10.9c}$$

$$A_3u = \tfrac{1}{4}[Au + iA(iu) + jA(ju) - kA(ku)], \tag{10.9d}$$

and where A_0 is linear, A_1 i-semilinear, A_2 j-semilinear, and A_3 k-semilinear; they are all additive. It is easy to verify that sums and scalar multiples of bounded additive operators are bounded and additive. Thus bounded additive operators form an R-module. It is also easy to verify that products of bounded additive operators are also bounded and additive. (For unbounded operators, as in the real and complex spaces, the domain is a proper subspace, and products of operators may or may not be defined.) That adjoints can be defined for linear operators was shown by Horwitz and Biedenharn,[1] and this has been demonstrated by an alternative method in Sec. V of this paper. For any τ-semilinear ($\tau = i, j$, or k) operator A, τA is linear, and this fact should enable one to extend the definition of the adjoint to τ-semilinear and thence to additive operators in general. There does not seem to be any difficulty in extending the definition of the norm of an operator from a complex space to a quaternionic one. Thus bounded additive operators form a normed $*$-algebra over $\mathbb{H}$.

It should be noted that it is not possible to define a *semilinear operator* analogous to the complex case in a quaternionic Hilbert space, that is, a semilinear operator A on a quaternionic Hilbert space $\mathscr{H}$ defined as an additive operator with the property that, for all $p \in \mathbb{H}$ and all $u \in \mathscr{H}$,

$$A(pu) = p^*Au \tag{10.10}$$

is inconsistent because

$$A(pqu) = p^*A(qu) = p^*q^*Au \tag{10.11}$$

and

$$A(pqu) = (pq)^*Au = q^*p^*Au, \tag{10.12}$$

and the two expressions are not necessarily equal because of the noncommutativity of the quaternions.

We believe that we have here the necessary groundwork for a systematic study of an operator algebra on a quaternionic Hilbert space. Such a study is in progress and will be reported in due course, but we hope that the reader gets a preview and a flavor of things to come.

ACKNOWLEDGMENTS

Much of the thinking behind this work was stimulated by some very pertinent questions asked by the referee of Ref. 3, and it is a pleasure to acknowledge this contribution. Numerous helpful suggestions for reducing the number of obscurities and inaccuracies from the referee of the present work are also gratefully acknowledged.

[1]L. P. Horwitz and L. C. Biedenharn, Ann. Phys. (NY) **157**, 432 (1984).
[2]S. L. Adler, Commun. Math. Phys. **104**, 611 (1986).
[3]C. S. Sharma and T. J. Coulson, J. Math. Phys. **28**, 1941 (1987).
[4]S.W. Hawking, in *General Relativity*, edited by S. W. Hawking and W. Israel (Cambridge U. P., Cambridge, 1979).
[5]C. Itzykson and J.-B. Zuber, *Quantum Field Theory* (McGraw-Hill, Singapore, 1985).

[6]J. Pian and C. S. Sharma, Nuovo Cimento B **94**, 193 (1986).
[7]J. Pian and C. S. Sharma, Int. J. Theor. Phys. **22**, 107 (1983).
[8]C. S. Sharma, Nuovo Cimento B **100**, 291 (1987).
[9]P.R. Halmos, *Finite-Dimensional Vector Spaces* (Van Nostrand, Princeton, 1958).
[10]R. O. Wells, *Differential Analysis on Complex Manifolds* (Springer, New York, 1980).
[11]S.-s. Chern, *Complex Manifolds Without Potential Theory* (Springer, New York, 1979).
[12]T. J. Coulson, Ph.D. thesis, London University, 1986.

A representation of SO($p+q$,$p+q$) on SO(p,q)

Marián Fecko
Department of Theoretical Physics, Comenius University, Mlynská dolina F2, 842 15 Bratislava, Czechoslovakia

(Received 7 August 1986; accepted for publication 22 December 1987)

A construction of a representation of SO($p+q$,$p+q$) by operators on SO($p+q$) is presented, connected with a relativistic top kinematics. In addition to a first-order differential operator part there is a multiplicative term containing a parameter λ in some of the generators. It is shown by the explicit evaluation of the corresponding Casimir invariants that the representation $(\lambda,...,\lambda)$, $\lambda = 0,\frac{1}{2},1,...$, is realized by this construction.

I. INTRODUCTION

Trying to formulate a relativistic theory of a top without enlarging its nonrelativistic configuration space [SO(3) group manifold], one is faced with a problem of realizing SO(3,1) on SO(3), which reduces (if standard generators of rotations are used for $\mathbf{s}$) to a realization of boost generators $\mathbf{N}$. The ansatz

$$N_i = B_{ik}s_k + b_i \tag{1}$$

leads to a system of equations

$$\begin{aligned} &\varepsilon_{jmn}B_{im}B_{nk} + \varepsilon_{kmn}B_{im}B_{jn} + \varepsilon_{imn}B_{jn}B_{mk} = \varepsilon_{ijk}, \\ &(B_{im}\varepsilon_{jmn} - B_{jm}\varepsilon_{imn})b_n = 0, \end{aligned} \tag{2}$$

for unknown quantities B_{ik} and b_i.[1] The general solution is

$$B_{ik} = \varepsilon_{ikj}n_j + \alpha n_i n_k, \quad \alpha,\beta \in \mathrm{C},$$

$$b_i = \beta n_i, \quad \mathbf{n} = \text{unit vector},$$

i.e.,

$$\mathbf{N} = -\mathbf{n}\times\mathbf{s} + \alpha\mathbf{n}(\mathbf{n}\cdot\mathbf{s}) + \beta\mathbf{n}. \tag{3}$$

It is shown in Ref. 1 that the free parameters α,β are connected with Casimir invariants $c_1^{\mathrm{L}} = -\mathbf{s}\cdot\mathbf{N}$, $c_0^{\mathrm{L}} = -(\mathbf{s}^2 - \mathbf{N}^2)$ of the Lorentz group, i.e., one can use them for the fixation of the spin of a top. (The case of spin $\frac{1}{2}$ was studied before in Ref. 2; it corresponds to $\alpha = 0$, $\beta = \frac{1}{2}$.)

In this paper we present a generalization of these results. We descibe a systematic construction of the generators of SO($p+q$,$p+q$) acting on functions on SO($p+q$). As in the above-mentioned case, nondifferential (multiplicative) terms containing a free parameter λ occur in some of the generators. By explicit evaluation of the corresponding Casimir invariants we identify the representation with $(\lambda,...,\lambda)$.

II. CONSTRUCTION OF THE GENERATORS

Let $\mathbf{e}_\alpha$, $\alpha = 1,2,...,(p+q) \equiv N$, form an orthonormal right-handed system of vectors in $M^{p,q}$, i.e.,

$$\eta^{\alpha\beta}e_{\alpha i}e_{\beta j} = \eta_{ij}, \tag{4}$$

$$\eta^{ij}e_{\alpha i}e_{\beta j} = \eta_{\alpha\beta}, \tag{5}$$

$$\varepsilon^{i_1\cdots i_N}e_{\alpha_1 i_1}\cdots e_{\alpha_N i_N} = (-1)^q\varepsilon_{\alpha_1\cdots\alpha_N}, \tag{6}$$

$$\varepsilon^{\alpha_1\cdots\alpha_N}e_{\alpha_1 i_1}\cdots e_{\alpha_N i_N} = (-1)^q\varepsilon_{i_1\cdots i_N} \tag{7}$$

hold, where

$$\eta^{ij} \equiv \mathrm{diag}(\underbrace{1,...,1}_{p}, \underbrace{-1,...,-1}_{q}), \tag{8}$$

$$\eta^{\alpha\beta} \equiv \mathrm{diag}(\underbrace{1,...,1}_{p}, \underbrace{-1,...,-1}_{q}), \tag{9}$$

and $e_{\alpha i}$ is the ith component of $\mathbf{e}_\alpha$. Generators of SO(p,q) can be represented by the antisymmetric tensor $S_{ij} = -S_{ji}$, $i,j = 1,...,(p+q)$, obeying

$$[S_{ij},S_{kl}] = -(n_{k[i}S_{j]l} + S_{k[j}n_{i]l}), \tag{10}$$

$$[S_{ij},e_{\alpha k}] = e_{\alpha[i}\eta_{j]k}, \tag{11}$$

where

$$A_{[kl]} \equiv A_{kl} - A_{lk}. \tag{12}$$

[The $e_{\alpha k}$ are to be expressed as functions of the coordinates on SO(p,q) and the S_{ij} are first-order differential operators with respect to the latter.]

One can construct additional operators on SO(p,q) now, combining $e_{\alpha i}$ and the S_{ij}:

$$S_{\alpha i} \equiv \eta^{kl}e_{\alpha k}S_{li} \equiv e_\alpha{}^k S_{ki}, \tag{13}$$

$$S_{\alpha\beta} \equiv e_\alpha{}^i e_\beta{}^j S_{ij} \equiv e_{\beta j}S_\alpha{}^j, \tag{14}$$

i.e., transforming successively the vector components of the S_{ij} to the scalar ones by means of "vielbein" $e_{\alpha i}$. Then using (10), (11) we obtain

$$[S_{ij},S_{\alpha k}] = S_{\alpha[i}\eta_{j]k}, \tag{15}$$

$$[S_{ij},S_{\alpha\beta}] = 0, \tag{16}$$

$$[S_{\alpha i},S_{\beta j}] = \eta_{\alpha\beta}S_{ij} - \eta_{ij}S_{\alpha\beta}, \tag{17}$$

$$[S_{\alpha i},S_{\beta\gamma}] = \eta_{\alpha[\gamma}S_{\beta]i}, \tag{18}$$

$$[S_{\alpha\beta},S_{\gamma\delta}] = \eta_{\gamma[\alpha}S_{\beta]\delta} + S_{\gamma[\beta}\eta_{\alpha]\delta}, \tag{19}$$

from which we deduce that the $S_{\alpha i}$ form $N \equiv (p+q)$ vectors and that the $S_{\alpha\beta}$ are scalars, both with respect to "right" rotations (generated by the S_{ij}). We can also change our point of view and classify objects according to their transformational properties with respect to "left" rotations (generated by the $S_{\alpha\beta}$). Then the S_{ij} are "scalars" and the $S_{\alpha i}$ form $N \equiv (p+q)$ "vectors" for fixed i on each.

This situation, for the special case $p = 3$, $q = 0$, is to some extent familiar from the theory of the nonrelativistic quantum-mechanical top,[3] where the projections on the laboratory as well as on the body axes of the quantities in question are used [including a change of sign in (19) in comparison with (10)]. "Mixed" operators $S_{\alpha i}$, however, are not discussed there at all.

It is not difficult to determine the algebra generated by S_{ij}, $S_{\alpha i}$, $S_{\alpha\beta}$. In order to do this we switch to a more compact

 0022-2488/88/051079-04$02.50

notation. Let us introduce an index $A \equiv (i,\alpha)$, $A = 1,...,2(p+q)$, $i = 1,...,N \equiv (p+q)$, $\alpha = N+1,...,2N$ (we changed the numeration of Greek indices). All generators form the components of a single object $S_{AB} = -S_{BA}$ now, and (10) and (15)–(19) read

$$[S_{AB},S_{CD}] = -(g_{C[A}S_{B]D} + S_{C[B}g_{A]D}), \tag{20}$$

where

$$g_{ij} \equiv \eta_{ij}, \tag{21}$$

$$g_{\alpha\beta} \equiv -\eta_{\alpha\beta}, \tag{22}$$

$$g_{\alpha i} \equiv g_{i\alpha} \equiv 0, \tag{23}$$

i.e.,

$$g_{AB} \equiv \mathrm{diag}(\underbrace{1,...,1}_{p}, \underbrace{-1,...,-1}_{q+p}, \underbrace{1,...,1}_{q}), \tag{24}$$

which reveals that the S_{AB} generate $\mathrm{SO}(p+q,p+q)$.

So far all generators S_{AB} were first-order differential operators on $\mathrm{SO}(p,q)$. Now we add a multiplicative term to $S_{\alpha i}$, introducing

$$\Sigma_{ij} \equiv S_{ij}, \tag{25}$$

$$\Sigma_{\alpha i} \equiv S_{\alpha i} + \lambda e_{\alpha i}, \tag{26}$$

$$\Sigma_{\alpha\beta} \equiv S_{\alpha\beta}, \tag{27}$$

where λ is a constant. Direct computation yields

$$[\Sigma_{AB},\Sigma_{CD}] = -(g_{C[A}\Sigma_{B]D} + \Sigma_{C[B}g_{A]D}) \tag{28}$$

independently of λ. That means that the Σ_{AB} form the generators of $\mathrm{SO}(p+q,p+q)$ as well.

Note that in the case $p=3$, $q=0$ discussed above we obtain—in addition to

$$s_i \equiv \tfrac{1}{2}\varepsilon_{ijk}S^{jk}, \tag{29}$$

$$s'_\alpha \equiv \tfrac{1}{2}\varepsilon_{\alpha\beta\gamma}S^{\beta\gamma}, \tag{30}$$

forming the $\mathrm{SO}(3)\otimes\mathrm{SO}(3)$ algebra of the laboratory and body projections of the angular momentum vector $\mathbf{s}$ of a top, respectively—three new vector operators Σ_α, $\alpha = 1,2,3$, where

$$\Sigma_{\alpha i} = e_\alpha{}^k S_{ki} + \lambda e_{\alpha i} = e_{\alpha k}\varepsilon_{kij}s_j + \lambda e_{\alpha i}$$
$$\equiv (-\mathbf{e}_\alpha \times \mathbf{s} + \lambda \mathbf{e}_\alpha)_i, \tag{31}$$

closing together with s_i and s'_α to $\mathrm{SO}(3,3)$ and thus offering a possibility of the relativization of the description of a top. We notice that (3) is just of the form of (31). [According to Ref. 1, for finite-dimensional representations of a Lorentz group one has to choose $\alpha = 0$, $\beta = 0,\frac{1}{2},1,...$ in (3).]

III. SPECIFICATION OF THE REPRESENTATION

In this section we evaluate Casimir invariants formed from (25)–(27), which enable us to specify the representation realized by this construction.

Several papers deal with the problem of the explicit form and eigenvalues of independent Casimir operators of classical groups (e.g., see Refs. 4–8). It was established in Ref. 7 that $N \equiv (p+q)$ invariants are to be evaluated in our case, viz., $(N-1)$ scalar operators C_n, $n = 2,4,...,2(N-1)$, and a pseudoscalar operator C'_N, where

$$C_n \equiv \Sigma^{A_1}{}_{A_2}\Sigma^{A_2}{}_{A_3}\cdots\Sigma^{A_n}{}_{A_1} \equiv (\Sigma^n)^A{}_A, \tag{32}$$

$$C'_N \equiv \varepsilon_{A_1B_1\cdots A_NB_N}\Sigma^{A_1B_1}\cdots\Sigma^{A_NB_N}. \tag{33}$$

A. Evaluation of C_n

Evaluation of C_n is based on the identity [specific for the construction (25)–(27)]

$$(\Sigma^2)^{AB} \equiv \Sigma^{AC}\Sigma_C{}^B = \lambda(\lambda+N-1)g^{AB} + (N-1)\Sigma^{AB}, \tag{34}$$

proved in Appendix A. This makes it possible to express an arbitrary "power" of Σ by Σ itself and a constant:

$$(\Sigma^n)^{AB} = a(n)g^{AB} + b(n)\Sigma^{AB} \tag{35}$$

(a and b can depend on λ and N in general as well). Multiplying (35) by $\Sigma_B{}^C$ we obtain recurrence relations

$$a(n+1) = a(2)b(n),$$
$$b(n+1) = a(n) + b(2)b(n), \tag{36}$$

or

$$\begin{pmatrix} a \\ b \end{pmatrix}_{n+1} = R\begin{pmatrix} a \\ b \end{pmatrix}_n, \tag{37}$$

where

$$R \equiv \begin{pmatrix} 0 & a(2) \\ 1 & b(2) \end{pmatrix} = \begin{pmatrix} 0 & \lambda(\lambda+N-1) \\ 1 & N-1 \end{pmatrix}, \tag{38}$$

so that

$$\begin{pmatrix} a \\ b \end{pmatrix}_n = R^{n-1}\begin{pmatrix} a \\ b \end{pmatrix}_1 = R^{n-1}\begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{39}$$

Evaluation of the necessary power of R gives

$$\begin{pmatrix} a \\ b \end{pmatrix}_n = \frac{1}{2\lambda+N-1}\begin{pmatrix} \lambda(\lambda+N-1)\{(\lambda+N-1)^{n-1} - (-\lambda)^{n-1}\} \\ (\lambda+N-1)^n - (-\lambda)^n \end{pmatrix}, \tag{40}$$

i.e.,

$$(\Sigma^n)^{AB} = \frac{\lambda(\lambda+N-1)\{(\lambda+N-1)^{n-1} - (-\lambda)^{n-1}\}}{2\lambda+N-1}g^{AB} + \frac{(\lambda+N-1)^n - (-\lambda)^n}{2\lambda+N-1}\Sigma^{AB}. \tag{41}$$

Then

$$C_n(\lambda,N) \equiv (\Sigma^n)^A{}_A = a(n)g^A_A = 2Na(n) = \frac{2N\lambda(\lambda+N-1)}{2\lambda+N-1}\{(\lambda+N-1)^{n-1} - (-\lambda)^{n-1}\}. \tag{42}$$

We observe that for each n, C_n is a number—not a differential operator, as is usually the case—when the harmonic analysis approach on corresponding homogeneous space is used. This means that the explicit construction of the functions on SO(p,q) in the space of which our representation of SO$(p+q,p+q)$ is realized reduces to solving the eigenvalue equations for the generators of a Cartan subalgebra (e.g., $\Sigma_{12},...,\Sigma_{2N-1,2N}$), i.e., only first-order differential equations are to be solved.

B. Evaluation of C'_N

For evaluation of $C'_N(\lambda,N)$ it is useful to realize that (41) reveals Σ_{AB} to be the only antisymmetric second-rank tensor (and g_{AB} the only symmetric one as well) available. That means then that also

$$\varepsilon_{A_1B_1A_2B_2\cdots A_NB_N}\Sigma^{A_2B_2}\cdots\Sigma^{A_nB_N}\sim\Sigma_{A_1B_1} \tag{43}$$

holds, which can be readily verified explicitly for not too large N, e.g., the coefficient of proportionality is 2 for $N=2$, $8(1+\lambda)$ for $N=3$, etc. (see Appendix B). Multiplication of both sides of (43) by $\Sigma^{A_1B_1}$ and taking into account (42) leads to the conclusion that C'_N does not contain differential operators, but reduces instead to the multiplication by a number, too (as was the case for all C_n). Thus we have to extract just this nondifferential part of (33), ignoring the differential terms completely (they cancel). We can introduce formally a symbol MULT, which when applied to any combined differential-multiplicative operator leaves its multiplicative part only. The above-mentioned conclusion can be written then as

$$C'_N(\lambda,N)=\text{MULT }C'_N(\lambda,N). \tag{44}$$

Nonvanishing components of a Levi-Civita tensor come from the cases where all indices are mutually different; in particular, an equal number of Latin and Greek indices should occur. The only distribution of indices which survives under the MULT symbol is one in which a pair of Latin and Greek indices stands on each Σ. [In the opposite case there is at least one Σ with both indices Latin, and it can be shifted to the right-hand side of the expression (33) giving a differential operator.] Taking into account two possibilities for the order of indices on each Σ ($i\alpha$ and αi) we can write

$$\text{MULT }C'_N(\lambda,N)=2^N\text{MULT }\varepsilon_{i_1\alpha_1\cdots i_N\alpha_N}\Sigma^{i_1\alpha_1}\cdots\Sigma^{i_N\alpha_N}, \tag{45}$$

and, using the result of Appendix C,

$$\begin{aligned}\text{MULT }&C'_N(\lambda,N)\\&=(-1)^{N(N-1)/2}2^N\varepsilon^{i_1\cdots i_N}\varepsilon^{\alpha_1\cdots\alpha_N}\\&\quad\times(S_{\alpha_1i_1}+\lambda e_{\alpha_1i_1})\cdots(S_{\alpha_Ni_N}+\lambda e_{\alpha_Ni_N}).\end{aligned} \tag{46}$$

With the help of the commutator

$$\begin{aligned}&\varepsilon^{i_1\cdots i_N}\varepsilon^{\alpha_1\cdots\alpha_N}[S_{\alpha_li_l},e_{\alpha_{p_1}i_{p_1}}\cdots e_{\alpha_{p_k}i_{p_k}}]\\&=k\varepsilon^{i_1\cdots i_N}\varepsilon^{\alpha_1\cdots\alpha_N}e_{\alpha_li_l}e_{\alpha_{p_1}i_{p_1}}\cdots e_{\alpha_{p_k}i_{p_k}}\end{aligned} \tag{47}$$

[for any $k,l,p_m=1,...,(N-1)$, $m=1,...,k$], we can perform the MULT operation in (46) explicitly and obtain a polynomial of order N in λ. Let us determine the coefficient standing by λ^r, $r=1,...,N$. It comes from all cases in which r "e" 's and $(N-r)$ "S" 's are chosen from the brackets in (46) ($e_{\alpha_Ni_N}$ should always come from the last one). Then the desired coefficient is

$$(-1)^{N(N-1)/2}2^NN!\sum_{p_1,...,p_{N-r-1}=1}^{N-1}\prod_{l=1}^{N-r-1}p_l \quad (\text{all } p_i \text{ different}), \tag{48}$$

where also [see (6),(7)]

$$\varepsilon^{i_1\cdots i_N}\varepsilon^{\alpha_1\cdots\alpha_N}e_{\alpha_1i_1}\cdots e_{\alpha_Mi_M}=N! \tag{49}$$

was used. One can compare it with the coefficient standing by λ^r in the expression

$$(\lambda+N-1)\cdots(\lambda+1)\lambda, \tag{50}$$

which, taking the numbers from $(N-r-1)$ brackets and λ from the rest, is just

$$\sum_{p_1,...,p_{N-r-1}=1}^{N-1}\prod_{l=1}^{N-r-1}p_l \quad (\text{all } p_i \text{ different}), \tag{51}$$

so that

$$\begin{aligned}&\text{MULT }\varepsilon^{i_1\cdots i_N}\varepsilon^{\alpha_1\cdots\alpha_N}(S_{\alpha_1i_1}+\lambda e_{\alpha_1i_1})\cdots(S_{\alpha_Ni_N}+\lambda e_{\alpha_Ni_N})\\&\quad=N!\lambda(\lambda+1)\cdots(\lambda+N-1),\end{aligned} \tag{52}$$

and

$$\begin{aligned}&C'_N(\lambda,N)\\&\quad=(-1)^{N(N-1)/2}2^NN!\lambda(\lambda+1)\cdots(\lambda+N-1).\end{aligned} \tag{53}$$

C. Comparison with the general results: Conclusion

Let us study $C'_N(\lambda,N)$ first. It is known[7] that its value is

$$\begin{aligned}C'_N(m_1,...m_N)&=(-1)^{N(N-1)/2}2^NN!(m_1+N-1)\\&\quad\times\cdots(m_{N-1}+1)m_N\end{aligned} \tag{54}$$

for the representation $(m_1,...,m_N)$. Our construction thus corresponds to the case $(\lambda,...,\lambda)$. However, the invariants C_n, $n=2,4,...,2(N-1)$, are to be compared, too. For the evaluation of C_n for $(\lambda,...,\lambda)$ we make use of the results of Ref. 8, where the generating function for Casimir invariants of all classical groups was derived. In the case of interest to us this function reads

$$\begin{aligned}G(z)=2N&\frac{1+(\lambda-N)z}{(1+\lambda z)(1-Nz)}\\&+\frac{2N\lambda(\lambda-1)}{(1+\lambda z)(1-Nz)(1-(\lambda+N-1)z)}\end{aligned} \tag{55}$$

and

$$G(z)=\sum_{n=0}^{\infty}C_nz^n, \tag{56}$$

so that

$$C_n=G^{(n)}(0)/n!. \tag{57}$$

The explicit calculation yields

$$C_n=\frac{2N\lambda(\lambda+N-1)}{2\lambda+N-1}\{(\lambda+N-1)^{n-1}-(-\lambda)^{n-1}\}, \tag{58}$$

in agreement with (42); that makes the identification with $(\lambda,...,\lambda)$ complete.

ACKNOWLEDGMENT

I would like to thank Professor M. Petráš for turning my attention to this problem and for permanent interest in the course of my work.

APPENDIX A: PROOF OF (34)

The proof is to be done for all four possible cases of pairs of indices, (ij), $(i\alpha)$, (αi), $(\alpha\beta)$. We restrict ourselves to the first case here; the rest can be done in the same way. We have

$$\begin{aligned}\Sigma_{iA}\Sigma^{A}{}_{j} &\equiv g^{AB}\Sigma_{iA}\Sigma_{Bj} = g^{kl}\Sigma_{lj} + g^{\alpha\beta}\Sigma_{i\alpha}\Sigma_{\beta j} = \eta^{kl}S_{ik}S_{lj} - \eta^{\alpha\beta}\Sigma_{i\alpha}\Sigma_{\beta j} = S_{ik}S^{k}{}_{j} + (e_{\alpha}{}^{m}S_{mi} + \lambda e_{\alpha i})(e^{\alpha n}S_{nj} + \lambda e^{\alpha}{}_{j})\\ &= S_{ik}S^{k}{}_{j} + e_{\alpha}{}^{m}e^{\alpha n}S_{mi}S_{nj} + e^{\alpha m}[S_{mi},e_{\alpha n}]S^{n}{}_{j} + \lambda e^{\alpha m}[S_{mi},e_{\alpha j}] + \lambda e_{\alpha}{}^{m}e^{\alpha}{}_{j}S_{mi} + \lambda e_{\alpha i}e^{\alpha n}S_{nj} + \lambda^{2}e_{\alpha i}e^{\alpha}{}_{j}\\ &= (N-1)S_{ij} + \lambda(\lambda+N-1)\eta_{ij} \equiv \lambda(\lambda+N-1)g_{ij} + (N-1)\Sigma_{ij}.\end{aligned}$$

(Notice that two "unpleasant" second-order terms cancel each other.) One should be careful when contracting Greek indices as to whether $g^{\alpha\beta} = -\eta^{\alpha\beta}$ [e.g., (45)] or $\eta^{\alpha\beta}$ is understood implicitly there.

APPENDIX B: (43) FOR $N=3$

We are to find the proportionality coefficient $\gamma(\lambda)$ in

$$\varepsilon_{ABCDEF}\Sigma^{CD}\Sigma^{EF} = \gamma(\lambda)\Sigma_{AB}.$$

Let us compute the $\alpha\beta$ component of the left-hand side:

$$\begin{aligned}\varepsilon_{\alpha\beta ABCD}\Sigma^{AB}\Sigma^{CD} &= 2\varepsilon_{\alpha\beta\gamma ijk}\Sigma^{\gamma i}\Sigma^{jk} + 2\varepsilon_{\alpha\beta jk\gamma i}\Sigma^{jk}\Sigma^{\gamma i} = -4\varepsilon_{ijk\alpha\beta\gamma}g^{\gamma A}\Sigma_{A}{}^{i}\Sigma_{jk} = 4\varepsilon_{ijk}\varepsilon_{\alpha\beta\gamma}(e^{\gamma i}S_{l}{}^{i} + \lambda_{l}{}^{\gamma i})S^{jk}\\ &= 4\epsilon_{ijk}e_{\alpha m}e_{\beta n}(-1)^{q}(\epsilon^{mnl}S_{l}{}^{i}S^{jk} + \lambda\varepsilon^{mni}S^{jk}) = 4(-1)^{q}\delta^{m}_{[i}\delta^{n}_{j]}e_{\alpha m}e_{\beta n}[s^{j},s^{i}] + 8\lambda e_{\alpha m}e_{\beta m}S^{mn}\\ &= 8(1+\lambda)S_{\alpha\beta} \equiv 8(1+\lambda)\Sigma_{\alpha\beta},\end{aligned}$$

where

$$\begin{aligned}&s_{i} \equiv \tfrac{1}{2}\varepsilon_{ijk}S^{jk}, \quad S_{ij} = (-1)^{q}\varepsilon_{ijk}s^{k},\\ &[s_{i},s_{j}] = -\varepsilon_{ijk}s^{k} = -(-1)^{q}S_{ij},\\ &\varepsilon_{\alpha\beta\gamma}e^{\gamma j} = (-1)^{q}\varepsilon^{mnj}e_{\alpha m}e_{\beta n},\\ &\varepsilon_{ijk}\varepsilon^{mnk} = (-1)^{q}\delta^{m}_{[i}\delta^{n}_{j]}\end{aligned}$$

was used, so that

$$\varepsilon_{ABCDEF}\Sigma^{CD}\Sigma^{EF} = 8(1+\lambda)\Sigma_{AB},$$

for $N=3$. Multiplying it by Σ^{AB} we obtain

$$\begin{aligned}C'_{3}(\lambda,3) &= -8(1+\lambda)\Sigma^{AB}\Sigma_{BA}\\ &\equiv -8(1+\lambda)C_{2}(\lambda,3)\\ &= -8(1+\lambda)(6\lambda(\lambda+2))\\ &\equiv -48\lambda(\lambda+1)(\lambda+2)\end{aligned}$$

in agreement with (53), for $N=3$. The same procedure is possible for arbitrary N.

APPENDIX C: PROOF OF (46)

We are to prove

$$\varepsilon_{i_1\alpha_1\cdots i_N\alpha_N} = (-1)^{N(N-1)/2}\varepsilon_{i_1\cdots i_N\alpha_1\cdots\alpha_N}.$$

Clearly

$$\varepsilon_{i_1\alpha_1\cdots i_N\alpha_N} = \varkappa(N)\varepsilon_{i_1\cdots i_N\alpha_1\cdots\alpha_N},$$

where $\varkappa(N)$ is 1 or (-1). Then

$$\begin{aligned}\varepsilon_{i_1\alpha_1\cdots i_N\alpha_N i_{N+1}\alpha_{N+1}} &= \varkappa(N)\varepsilon_{i_1\cdots i_N\alpha_1\cdots\alpha_N i_{N+1}\alpha_{N+1}}\\ &= (-1)^{N}\varkappa(N)\varepsilon_{i_1\cdots i_{N+1}\alpha_1\cdots\alpha_{N+1}}\\ &= \varkappa(N+1)\varepsilon_{i_1\cdots i_{N+1}\alpha_1\cdots\alpha_{N+1}},\end{aligned}$$

so that

$$\begin{aligned}\varkappa(N+1) &= (-1)^{N}\varkappa(N) = (-1)^{N}(-1)^{N-1}\varkappa(N-1)\\ &= \cdots = \varkappa(1)\prod_{k=0}^{N-1}(-1)^{N-k}\\ &= \prod_{k=1}^{N}(-1)^{k} = (-1)^{\Sigma_{k=1}^{N}k} = (-1)^{N(N+1)/2},\end{aligned}$$

and

$$\varkappa(N) = (-1)^{N(N-1)/2}.$$

[1] M. Fecko, *Proceedings of the 8th Conference of Czechoslovac Physicists*, Bratislava, 1985 (JČSMF, Prague, 1985).

[2] M. Petráš, unpublished.

[3] L. D. Landau, and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, Oxford, 1958).

[4] G. Racah, *Group Theory and Spectroscopy* (Princeton U.P., Princeton, NJ, 1951).

[5] L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).

[6] M. Umezawa, Nucl. Phys. **48**, 111 (1963); **53**, 54 (1964); **57**, 65 (1964).

[7] A. M. Perelomov and V. S. Popov, Sov. J. Nucl. Phys. **3**, 1127 (1966).

[8] V. S. Popov and A. M. Perelomov, Sov. J. Nucl. Phys. **7**, 460 (1968).

Locally operating realizations of nonconnected transformation Lie groups

Oscar García-Prada[a)]
Departamento de Física Teórica, Facultad de Ciencias, Universidad de Valladolid, 47011 Valladolid, Spain
Mariano A. del Olmo[b)] and Mariano Santander[b)]
Centre de Recherches Mathématiques, Université de Montréal, Montréal H3C 3J7, Canada

(Received 15 July 1986; accepted for publication 20 October 1987)

A systematic study of locally operating multiplier realizations of nonconnected Lie groups of transformations is presented that generalizes previous results on connected groups. The semilinear locally operating multiplier realizations of a nonconnected group G are those obtained through an induction process from the finite-dimensional semilinear representations of a given subgroup of a representation group $\overline{G}$ for G.

I. INTRODUCTION

The concept of locally operating realization of Lie groups of transformations was developed some years ago by Hoogland[1,2] who also showed its physical relevance in quantum mechanics. From a physical point of view, the relevant realizations of a symmetry group G are not the linear ones, but also the projective representations.[3,4] If one considers such representations to be locally operating, one obtains locally operating multiplier representations of G, called hereafter simply locally operating realizations.

The study of these realizations was first made by Hoogland,[1,2] who studied the cases of both connected and nonconnected groups of transformations, the latter allowing for the appearance of semilinear operators and semilinear multiplier locally operating representations.[2]

In recent papers[5–9] Mackey's theory of induced representations[10] has been pointed out as the appropriate mathematical framework for the study of locally operating realizations (LOR's). That theory was worked out for a connected Lie group, and gives both a complete characterization of locally operating realizations as induced representations, as well as an explicit construction method of these realizations. This method makes use of a new group $\overline{G}$, called the local splitting group, such that all LOR's of G appear through the linear ones of $\overline{G}$.

From the preceding results one may expect the same construction being valid for nonconnected Lie groups. The purpose of this work is to report that this is actually the case, and in this paper we give a generalization to the nonconnected case of the theory developed previously for connected groups. Although most of the results seem an almost verbatim transcription of the results for the connected case, the explicit final form of the LOR's of a concrete nonconnected group depends on what group elements are being represented linearly or antilinearly, and also on the structure of the group as an extension of its connected component of the identity G_0 by its components group, $\pi_0(G)$. In most cases, including some physically relevant ones which we include as examples, that structure is a semidirect product which greatly simplifies the final results. In these cases it would be also possible to obtain the pseudoequivalence classes of locally operating realizations of the complete group in a "direct" way, perhaps at the price of a greater amount of computations as Wigner[11] pointed out for the relativistic case. We remark that the theory developed in this paper holds no matter what structure of G, whether as a semidirect product or as some essential extension of G_0 by $\pi_0(G)$.

The paper is organized as follows: Sec. II is devoted to a short review of the main results about locally operating realizations in the connected case.[8,9] In Sec. III we state the conditions on some group $\overline{G}$ that allows one to obtain the semilinear locally operating multiplier representations of G as semilinear representations of $\overline{G}$. Such a group will be called a representation group. Here we give all the relevant results about the equivalence of the realizations so obtained as well as the method to explicitly obtain a complete set of representatives of the pseudoequivalence classes. In Sec. IV we present some examples.

II. LOCALLY OPERATING REALIZATIONS

We give here a short review of the main concepts of the theory of locally operating representations for connected Lie groups. A more detailed account may be found in Refs. 8 and 9.

Let G be a connected Lie group, acting transitively on a differentiable manifold X, and let Γ be the isotopy group of a fixed point $x_0 \in X$. The subgroup Γ is closed and the homogeneous space G/Γ can be endowed with a differentiable structure such that G/Γ is diffeomorphic to X. A locally operating (multiplier) realization (hereafter simply LOR) of G is a Borel multiplier representation U of G, acting in the space of vector-valued complex functions f: $X \to \mathbf{C}^n$ in the following way:

$$[U(g)f](gx) = A(g,x)f(x)\,, \tag{2.1}$$

where A is a matrix-valued Borel function A: $G \times X \to \mathrm{GL}(n,\mathbf{C})$ called a gauge matrix. The relevant concept of equivalence for LOR is that of gauge pseudoequivalence, which is finer than the usual equivalence. Two LOR's U and U' of G are said to be pseudoequivalent if there exists a Borel function λ: $G \to \mathrm{U}(1)$ and an invertible linear operator

[a)] Present address: Department of Mathematics, Wiess School of Natural Sciences, Rice University, Houston, Texas 77251.
[b)] Permanent address: Departamento de Física Teórica, Facultad de Ciencias, Universidad de Valladolid, 47011 Valladolid, Spain.

τ acting locally in the representation space [i.e., $(\tau f)(x) = S(x)f(x)$ with S a nonsingular matrix-valued Borel function] such that $\forall g \in G$,

$$U'(g) = \lambda(g)\tau U(g)\tau^{-1}. \quad (2.2)$$

The corresponding gauge matrices are related by

$$A'(g,x) = \lambda(g)S(gx)A(g,x)S^{-1}(x), \quad \forall g \in G, \quad \forall x \in X. \quad (2.3)$$

Notice that gauge pseudoequivalence is characterized by the requirement that the intertwining operator τ acts locally, i.e., as matrix multiplication with an x-dependent matrix. Thus gauge equivalence implies ordinary equivalence, but the converse is not true.

In the linear case, the locally operating *linear* (no multiplier) representations of a group G (hereafter LOLR) are induced from the finite-dimensional linear representations of the isotopy group Γ of the action of G on X. The gauge matrix associated to a LOLR of G is given by

$$A(g,x) = \sigma(s^{-1}(gx)gs(x)), \quad \forall g \in G, \quad x \in X, \quad (2.4)$$

where σ is a finite-dimensional representation of Γ and s is a normalized Borel secton s: $G/\Gamma \approx X \to G$. Moreover, each equivalence class of finite-dimensional representations of Γ gives a gauge equivalence class of LOLR's of G.

Turning now to the general case of *multiplier* realizations, the first step is to reduce the multiplier problem to a linear one. This is made using an auxiliary group $\overline{G}$, known as a splitting group[12,13] for G, such that any LOR of G can be lifted to a *linear* LOR of $\overline{G}$. For the general case of all (not necessary local) representations, such a group $\overline{G}$ can be found as some extension of G by the dual $\widehat{H^2(G,U(1))}$ of the second cohomology group of G [13]. It was shown in Ref. 8 that for LOR's this linearization trick does also work with a small change. A group which linearizes *all* LOR's of G is said to be a local splitting group for G, and if it is in some sense minimal one calls it a local representation group, simply denoted $\overline{G}$. In fact, not all factor systems of G could appear in local representations, but only those corresponding to a subgroup of $H^2(G,\mathrm{U}(1))$, called $H^2_{\mathrm{loc}}(G,\mathrm{U}(1))$. As one can expect, $\overline{G}$ is a topological extension of G by the dual $\widehat{H^2_{\mathrm{loc}}(G,\mathrm{U}(1))}$, i.e., we have an exact sequence

$$1 \to \widehat{H^2_{\mathrm{loc}}(G,\mathrm{U}(1))} \to \overline{G} \xrightarrow{p} G \to 1,$$

where p is an epimorphism.

The characterization of $H^2_{\mathrm{loc}}(G,\mathrm{U}(1))$ is given in Ref. 7. Its elements are those classes of factor systems $[\omega] \in H^2(G,\mathrm{U}(1))$ whose restriction to $\Gamma \times \Gamma$ lives in the maximal compact group of $H^2(\Gamma,\mathrm{U}(1))$.

The explicit construction of the local representation group is given in Ref. 8. The action of G on X induces an action of $\overline{G}$ into X via the epimorphism p. Then, if we denote by $\overline{\Gamma}$ the isotropy group of a fixed point $x_0 \in X$, we have $\overline{\Gamma} = p^{-1}(\Gamma)$. Once $\overline{G}$ is known, the LOR U of G are obtained through the split LOR's R of $\overline{G}$ [split means $R(\ker p) \subset \mathrm{U}(1)$], according to the relation $U(g) = (R \circ \rho)(g)$, where ρ is some normalized Borel section, ρ: $G \to \overline{G}$. And finally, these LOR's of $\overline{G}$ are just those which can be induced from the finite-dimensional linear representations of $\overline{\Gamma}$ mapping $\ker p$ into $\mathrm{U}(1)$, and called split representations, too.

The most important result is the one-to-one correspondence between the pseudoequivalence classes of LOR of G and the superequivalence classes of finite-dimensional split linear representations of $\overline{\Gamma}$.

We here recall that two pseudoequivalent representations of $\overline{\Gamma}$ are said to be superequivalent if the one-dimensional homomorphism of $\overline{\Gamma}$ on $\mathrm{U}(1)$ realizing their pseudoequivalence [see (2.2)] can be extended to a homomorphism of $\overline{G}$ on $\mathrm{U}(1)$.

III. LOCALLY OPERATING REALIZATIONS OF NONCONNECTED LIE GROUPS

A. Definitions

From now on we will assume that G is a nonconnected Lie group acting transitively on a differentiable manifold X. Let Γ be the isotopy group of a fixed point $x_0 \in X$. We can identify X with G/Γ. We shall be interested in *semilinear* realizations R of (G,H), that is, realizations of G on the group of linear and antilinear operators in a Hilbert space such that H is a subgroup of index 1 or 2 in G, and $R(g)$ is a linear or antilinear operator according as $g \in H$ or $g \in G - H$.

In this way the natural definition of a semilinear locally operating multiplier realization is the following.

Definition 1: A semilinear locally operating multiplier realization (SLOR) of (G,H) is a Borel semilinear multiplier realization of (G,H) which acts in a representation space of functions f: $X \to \mathbf{C}^n$ as

$$[U(g)f](gx) = A(g,x)f^g(x), \quad (3.1)$$

where $f^g(x) = f(x)$ or $f^*(x)$ according as $g \in H$ or $g \in G - H$, and $A(g,x)$ is a nonsingular matrix-valued function A: $G \times X \to \mathrm{GL}(n,\mathrm{C})$, which verifies the following relation:

$$A(g',gx)A^{g'}(g,x) = \omega(g',g)A(g'g,x), \quad (3.2)$$

$\omega(g',g)$ being a factor system of G relative to the action of G on $\mathrm{U}(1)$ given by identity if $g \in H$ or complex conjugation if $g \in G - H$. Hence $\omega \in Z^2{}_{*H}(G,\mathrm{U}(1))$, the symbol $*H$ in the Z^2 denoting this action.

The definition of gauge pseudoequivalence is carried out as in the linear case; we could here distinguish between linear and semilinear pseudoequivalence according to the linear or antilinear character of the operator τ.

It is very easy to show that if U and U' are gauge pseudoequivalent, then their gauge matrices are related by

$$A'(g,x) = \lambda(g)S(gx)A(g,x)S^{-1g}(x). \quad (3.3)$$

B. Semilinear (no multiplier) locally operating representations of (G,H)

These representations are a particular case of SLOR of (G,H) when their factor systems are equal to 1, and hence there are no multipliers.

The proof that (nonmultiplier) semilinear LOR's of (G,H) are just those representations of G induced from the semilinear finite dimensional representations of $(\Gamma,\Gamma \cap H)$ is similar to the one presented in Ref. 5 for the connected case. For this reason we do not give here the demonstration nor the explicit construction. Nevertheless, there are some minor differences due to the fact we are working now with semilinear representations. The interested reader must find no trouble in carrying out his/her own proof, following Ref. 5 and paying attention to the specific differences which arise from the semilinear operators.

The main result can be summarized in the following theorem.

Theorem 1: A complete set of gauge equivalence classes of SLOR's of (G,H) is obtained by induction from a complete set of equivalent classes of finite-dimensional semilinear representations of $(\Gamma,\Gamma\cap H)$. If σ is a representative of an equivalence class of finite-dimensional semilinear representations of $(\Gamma,\Gamma\cap H)$, the associated gauge matrix of the induced representation of (G,H) is given by

$$A(g,x)=\tilde{\sigma}(s^{-1}(gx)gs(x)), \tag{3.4}$$

where $\tilde{\sigma}(s^{-1}(gx)gs(x))$ is the "matrix part" of the representation σ [i.e., $\sigma(\gamma)=\tilde{\sigma}(\gamma)$ or $\tilde{\sigma}(\gamma)\mathbf{K}$ according as $\gamma\in\Gamma\cap H$ or $\gamma\notin\Gamma\cap H$, $\mathbf{K}$ being the complex conjugation in $\mathbf{C}^n$], and s is a normalized Borel section s: $G/\Gamma\rightarrow G$ verifying $s^{-1}(gx)gs(x)\in\Gamma\cap H$ for every $\gamma\in H$.

Not all sections s: $G/\Gamma\rightarrow G$ satisfy this last condition, but in the cases where the connected component of the identity G_0 acts transitively on X, such a section with $s(G/\Gamma)\subset G_0$ does always exist. Unlike the connected case, the election of a section has to satisfy that additional requirement.

C. Local representation groups

As previously remarked, for any group (G,H), it is always possible to construct a splitting group $(\overline{G},\overline{H})$ in such a way that all semilinear multiplier representation of (G,H) can be obtained from the semilinear ones of $(\overline{G},\overline{H})$; and as pointed out in Sec. III B, each SLOR $\overline{G}$ gives rise to a (multiplier) SLOR of G, due to the relation between the action of G and $\overline{G}$ into X. Hence although we could take for $\overline{G}$ any representation group, we notice that not all, but only those classes in $H^2{}_{*H}(G,\mathrm{U}(1))$ for which the restriction to Γ appears as a class of factor systems of some finite-dimensional semilinear representation of Γ could appear as factor systems of LOR of G.

Therefore, for our purposes, it suffices to take for $\overline{G}$ a local representation group, i.e., a topological extension of G by the dual of the subgroup of $H^2{}_{*H}(G,\mathrm{U}(1))$ whose classes appear in local representations. It does not seem easy to obtain an explicit characterization of this subgroup $H^2_{\mathrm{loc}*H}(G,U(1))$ as the one given before for connected groups. Nevertheless, in all the cases we study in this paper, the subgroup $H^2_{\mathrm{loc}*H}(G,\mathrm{U}(1))$ can always be easily found after its defining property, instead of in terms of some characterization. It is easy to prove that $H^2_{\mathrm{loc}*H}(G,\mathrm{U}(1))$ is a closed subgroup of $H^2{}_{*H}(G,\mathrm{U}(1))$ and hence a Lie group, too, in case $H^2{}_{*H}(G,\mathrm{U}(1))$ is a Lie group.

The construction of the local representation group follows closely the pattern of the corresponding one for the representation group in the general case.[13] We must recall that $\overline{G}$ is a noncentral extension, the action of G on $H^2_{\mathrm{loc}*H}(G,\mathrm{U}(1))\widehat{\ }$ as the identity or the inversion according as $g\in H$ or $g\in G-H$; and the factor system characterizing the extension, say, $W(g,h)$, is found through a homomorphic section s: $H^2_{\mathrm{loc}*H}(G,\mathrm{U}(1))\rightarrow Z^2_{\mathrm{loc}*H}(G,\mathrm{U}(1))$ just as in the nonlocal case. Note that if R is a split representation of $\overline{G}$, then $R(W(g,h),1)$ is the factor system of the multiplier representation of G obtained by projection of R.

D. Lifting of semilinear locally operating realizations of (*G*,*H*)

In this subsection we will assume the existence of a local representation group $(\overline{G},\overline{H})$ for (G,H). The action of G on X induces an action of $\overline{G}$ on X by $\overline{g}x=p(\overline{g})x=gx$, with p: $\overline{G}\rightarrow G$ the canonical epimorphism. The isotopy group of a fixed point x_0 of X will be $\overline{\Gamma}=p^{-1}(\Gamma)$. Hence we have the following proposition.

Proposition 1: (i) For each normalized Borel section ρ for p, i.e., $p\circ\rho=id_G$, and for each split SLOR R of $(\overline{G},\overline{H})$ there exists a SLOR U of (G,H) given by $U(g)=(R\circ\rho)(g)$, $\forall g\in G$.

(ii) Conversely, for each SLOR U of (G,H) there is another SLOR U' of (G,H), gauge pseudoequivalent to U and such that U' can be lifted to a split SLOR of $(\overline{G},\overline{H})$.

The proof reduces to some calculations taking into account the connection $\overline{A}(\rho(g),x)=A(g,x)$ between the gauge matrices corresponding to R and $U=R\circ\rho$, the relation $U'(g)=\lambda(g)U(g)$, and the decomposition of the elements of $\overline{G}$ as $\overline{g}=a\rho(g)$, with $a\in\ker p$ and $g\in G$.

It can be also easily verified that the finite-dimensional semilinear representations of $(\overline{\Gamma},\overline{\Gamma}\cap\overline{H})$ inducing split SLOR's of $(\overline{G},\overline{H})$ are those mapping $\ker p$ on $\mathrm{U}(1)$. We will also call them split representations.

From all these results we can obtain the following theorem.

Theorem 2: If $(\overline{G},\overline{H})$ is a local representation group for (G,H), for each normalized Borel section ρ of G on $\overline{G}$ there exists a one-to-one correspondence of the gauge pseudoequivalence classes of SLOR's of (G,H) with the gauge pseudoequivalence classes of split SLOR of $(\overline{G},\overline{H})$.

As we know from Sec. III A the SLOR's of $(\overline{G},\overline{H})$ are induced from the finite-dimensional split semilinear representations of $(\overline{\Gamma},\overline{\Gamma}\cap\overline{H})$. The relevant equivalence for the representations of splitting groups is the pseudoequivalence.[13] However, pseudoequivalent representations of $(\overline{\Gamma},\overline{\Gamma}\cap\overline{H})$ do not induce, in general, gauge pseudoequivalent SLOR's of $(\overline{G},\overline{H})$. The correct answer to this problem is given by the following theorem.

Theorem 3: Two split pseudoequivalent finite-dimensional semilinear representations of $(\overline{\Gamma},\overline{\Gamma}\cap\overline{H})$ induce gauge pseudoequivalent split SLOR's of $(\overline{G},\overline{H})$ if and only if the element $\Lambda\in Z^1{}_{*\overline{\Gamma}\cap\overline{H}}(\overline{\Gamma},\mathrm{U}(1))$ defining the pseudoequivalence of the representations of $(\overline{\Gamma},\overline{\Gamma}\cap\overline{H})$ can be extended to an element $\lambda\in Z^1{}_{*\overline{H}}(\overline{G},\mathrm{U}(1))$.

Proof: Let R and R' be two gauge pseudoequivalent SLOR's of $(\overline{G},\overline{H})$. Then there exists a mapping λ: $\overline{G}\rightarrow\mathrm{U}(1)$ satisfying (2.2) and the fact that R and R' are semilinear representations implies that λ is a crossed homomorphism, i.e., $\lambda\in Z^1{}_{*\overline{H}}(\overline{G},\mathrm{U}(1))$. The gauge matrices R and R' restricted to $\overline{\Gamma}\times\{x_0\}$ are pseudoequivalent matrix representations of $\overline{\Gamma}$, and their pseudoequivalence is carried by $\lambda|_{\overline{\Gamma}}$. Conversely, if we have two pseudoequivalent semilinear representations of $\overline{\Gamma}$, σ and σ', there exists $\Lambda\in Z^1{}_{*\overline{\Gamma}\cap\overline{H}}(\overline{\Gamma},\mathrm{U}(1))$ which determines this pseudoequivalence. According to the theory of induced representations, the representations R and

R' of $\overline{G}$ induced by σ and σ' are pseudoequivalent [see formulas (2.2) and (3.4)]. However, R and R' will be only gauge pseudoequivalent [see (3.3)] if the term $\Lambda(\overline{s}^{-1}(\overline{g}x)\overline{g}\,\overline{s}(x))$ can be split as

$$\Lambda(\overline{s}^{-1}(\overline{g}x)\overline{g}\,\overline{s}(x))=\Lambda(\overline{s}^{-1}(\overline{g}x))\Lambda(\overline{g})\Lambda(\overline{s}(x)),$$

that is, if Λ can be extended to a crossed homomorphism of $\overline{G}$. Here $\overline{s}$ is a Borel section of X on $\overline{G}$ satisfying the conditions stated after Theorem 1.

This theorem motivates the following definition which generalizes the corresponding one of the connected case.

Definition 2: Two split finite-dimensional semilinear representations of $(\overline{\Gamma},\overline{\Gamma}\cap\overline{H})$ are called superequivalent if they are pseudoequivalent and this pseudoequivalence can be realized by a crossed homomorphism of $\overline{\Gamma}$ on U(1) that can be extended to a crossed homomorphism of $\overline{G}$ on U(1).

Representatives σ of all the classes of superequivalence are obtained by $\hat{\sigma}=\lambda\circ\sigma$, where λ runs through a complete set of equivalence classes of crossed homomorphisms of $\overline{\Gamma}$ on U(1) modulo those that can be extended to $\overline{G}$ and σ runs through the pseudoequivalence classes of finite-dimensional split semilinear representations of $(\overline{\Gamma},\overline{\Gamma}\cap\overline{H})$, respectively.

The following theorem gives the explicit construction of the gauge equivalence classes of SLOR's of $(\overline{G},\overline{H})$ and summarizes the above results.

Theorem 4: (i) A complete set of gauge pseudoequivalence classes of split SLOR's of $(\overline{G},\overline{H})$ is obtained by induction from a complete set of superequivalence classes of the split semilinear representations of $(\overline{\Gamma},\overline{\Gamma}\cap\overline{H})$. If $\hat{\sigma}$ denotes a representative of each of such classes, then the induced SLOR of $(\overline{G},\overline{H})$ has a gauge matrix given by $\overline{A}(\overline{g},x)=\tilde{\sigma}(\overline{s}^{-1}(\overline{g}x)\overline{g}\,\overline{s}(x))$, where $\tilde{\sigma}$ is the matrix part of the representation $\hat{\sigma}$ of $\overline{\Gamma}$ and s is a normalized Borel section of X on $\overline{G}$.

(ii) A complete set of gauge pseudoequivalence classes of SLOR's of (G,H) with representatives $U=R\circ\rho$, where ρ is a normalized Borel section of G on $\overline{G}$, is obtained when R runs through a complete set of representatives of SLOR's of $(\overline{G},\overline{H})$. The gauge matrix of U is $A(g,x)=\overline{A}(\rho(g),x)$, with $\overline{A}$ the gauge matrix of R. The factor system $\omega\in Z^2{}_{*H}(G,\mathrm{U}(1))$ associated to U is $\omega=R\circ W_\rho$, where W_ρ is the factor system of the extension of G via the section ρ, i.e.,

$$W_\rho(g',g)=\rho(g')\rho(g)\rho^{-1}(g'g).$^{[13]}$$

A particular choice of the sections ρ and s may lead to simpler expressions. For example, we can choose $\rho(g)=(1,g)$ and $\overline{s}(x)=(1,s(x))$, where s is a normalized Borel section of X on G verifying $s(X)\subset H$. Thus the gauge matrix of U is written as

$$A(g,x)=\hat{\sigma}((1,s(gx))^{-1}(1,g)(1,s(x))). \tag{3.5}$$

IV. EXAMPLES

A. The complete Euclidean group E(2)

Let $E_0(2)$ be the proper Euclidean group, which acts on $\mathbf{R}^2$ in the standard way. We consider the group E(2) obtained by adjoining to $E_0(2)$ the line reflection P,

$$P:\begin{pmatrix}x^1\\x^2\end{pmatrix}\to\begin{pmatrix}x^1\\-x^2\end{pmatrix}.$$

As an abstract group, E(2) is a semidirect product of $E_0(2)$ and Z_2, the action of Z_2 into $E_0(2)$ being given by

$$P:(a_1,a_2,\phi)\to(a_1,-a_2,-\phi).$$

The LOR (up to pseudoequivalence) of $E_0(2)$ are parametrized by $\beta\in\mathbf{R}$, corresponding to $H^2{}_{\mathrm{loc}}(E_0(2),\mathrm{U}(1))\cong\mathbf{R}$, and are given[9] by

$$\{U_\beta(g)f\}(g\mathbf{x})=\{\exp[-i(\beta/2)(\mathbf{a}\wedge\mathbf{x}^\phi)_3]\}f(\mathbf{x}).$$

When the SLOR's of E(2) are considered we have two choices for P: $U(P)$ is either a linear or an antilinear operator.

(a) We take **P** to be represented as a *linear* (unitary) *operator*, and then $H=\mathrm{E}(2)$.

The first task is to compute the group $H^2{}_{*H}(\mathrm{E}(2),\mathrm{U}(1))$. This is carried out making use of the fact that every factor system of a semidirect product $G_0\odot Y$ is equivalent to one of the form

$$\omega(g',\alpha';g,\alpha)=\omega^{G_0}(g',g^{\alpha'})\omega^V(\alpha',\alpha)\Lambda(g,\alpha'), \tag{4.1}$$

where ω^{G_0},ω^V are factor systems of G_0 and V, and Λ has to satisfy some equations (see the Appendix). In fact, for our case $\omega^{G_0}(g',g)$ can be taken as $\exp(-i\beta/2(\mathbf{a}\wedge\mathbf{a}^{\phi'})_3)$. On the other hand $H^2{}_{*Z}(Z_2,\mathrm{U}(1))$ is trivial, and as ω^V we can take the trivial factor system. Then the equations for Λ have no solution unless $\beta=0$. That is, $H^2{}_{*H}(\mathrm{E}(2)\mathrm{U}(1))=\{1\}$.

Hence if $U(P)$ is a linear operator there are *no* LOR's of the complete Euclidean group whose restriction to the proper group is a realization with $\beta\neq0$.

(b) We now take **P** to be represented as an *antilinear operator*, and hence $H=E_0(2)$. In this case things are very different, and for the group $H^2{}_{*\mathrm{E}(2)}(\mathrm{E}(2),\mathrm{U}(1))$ we get $H^2{}_{*\mathrm{E}}(\mathrm{E}(2),\mathrm{U}(1))=\mathbf{R}\otimes\mathbf{Z}_2$, a generic element being denoted $[\beta,m]$, $\beta\in\mathbf{R}$, $m\in\{1,-1\}$. A lifting of a factor system in the class $[\beta,m]$ is given by

$$\omega_{\beta,m}(g',\alpha';g,\alpha)=\omega_\beta^{E_0(2)}(g',g^{\alpha'})\omega_m^{\mathbf{Z}_2}(\alpha',\alpha) \tag{4.2}$$

[here $\omega_m^{\mathbf{Z}_2}(P,P)=m$; see Ref. 13 (Sec. IX B) for details].

The representation group for E(2) is now an extension of the representation group for $E_0(2)$ by $\overline{\mathbf{Z}}_2$, with law

$$(\theta',g',\gamma',\alpha')(\theta,g,\gamma,\alpha)=(\theta'+\theta^{\alpha'}+\omega_\beta^{E_0(2)}(g',g^{\alpha'}),g'g^{\alpha'},\gamma'\gamma\omega_m^{\mathbf{Z}_2}(\alpha',\alpha),\alpha'\alpha) \tag{4.3}$$

with $\theta\in\mathbf{R}$, $g\in\mathrm{E}(2)$, $\gamma\in\{1,-1\}$, $\alpha\in\{1,P\}$, where the group $\overline{\mathbf{Z}}_2$ is isomorphic to Z_4, generated by $(1,P)$ with $(1,P)^2=(-1,1)$, and $\theta^P=-\theta$.

The group acts via the projection $(\theta,g,\gamma,\alpha)\to(g,\alpha)$ on the plane and the isotopy group $\overline{\Gamma}$ is the set $\{(\theta,(0,0,\phi),\gamma,\alpha)\}$ which has a semidirect product structure $\{\mathbf{R}\otimes\mathrm{SO}(2)\}\odot\mathbf{Z}_4$ with $\mathbf{Z}_4$ acting on $\mathbf{R}\otimes\mathrm{SO}(2)$ via the projection $(\gamma,\alpha)\to\alpha$.

As follows from the theory given in Sec. III we have to compute the sets of one-dimensional semilinear representations (crossed homomorphisms) of $\overline{\Gamma}$ and $\overline{\mathrm{E}}(2)$ on U(1). Both are easily found using the Wigner procedure[14] starting from the linear representations of $\overline{\Gamma}_+$ and $\overline{\mathrm{E}}(2)_+$ (the symbol + indicates the subgroup represented by linear operators, i.e., $\overline{H}$). The result is that the classes of one-dimensional semilinear representations of $\overline{\Gamma}$ modulo those which are extended to $\overline{\mathrm{E}}(2)$ are parametrized by β. Hence the general form of a SLOR of the complete Euclidean group induced by a one-dimensional representation of $\overline{\Gamma}$ is given by

$$(U(g,\alpha)f)((g,\alpha)\bar{x}) = \exp(-i(\beta/2)(\mathbf{a}\wedge\mathbf{x}^{(\phi,\alpha)})_3)\Delta(\alpha)f(\mathbf{x}), \tag{4.4}$$

where $\Delta(P)=\mathbf{K}$. In the calculation we have taken the section $\bar{s}$: $X\to\overline{\mathrm{E}}(2)$ given by $\bar{s}(\mathbf{x})=(0,\mathbf{x},0,1,1)$.

B. Kinematical groups

1. Galilei (2+1) group

Let us now consider the complete Galilei group G in two space dimensions, obtained by adjoining to the proper Galilei $(2+1)$ group G_0 the line reflection P and the time reversal T,

$$P: \begin{pmatrix} x_1 \\ x_2 \\ t \end{pmatrix} \to \begin{pmatrix} x_1 \\ -x_2 \\ t \end{pmatrix}, \quad T: \begin{pmatrix} x_1 \\ x_2 \\ t \end{pmatrix} \to \begin{pmatrix} x_1 \\ x_2 \\ -t \end{pmatrix}.$$

The group is a semidirect product of the connected Galilei group $G_0=\{(b,a_1,a_2,v_1,v_2,\phi)\}$ by a group of reflections isomorphic to the Klein Vierergruppe V, acting on G_0 as

$$P:\ (b,a_1,a_2,v_1,v_2,\phi)\to(b,a_1,-a_2,v_1,-v_2,-\phi),$$
$$T:\ (b,a_1,a_2,v_1,v_2,\phi)\to(-b,a_1,a_2,-v_1,-v_2,\phi),$$

and its elements will be naturally denoted (g,α), with $g\in G_0$ and $\alpha\in V$.

In this case it is known[15] that $H^2(G_0,\mathrm{U}(1))\cong\mathbf{R}^2$, an element $[M,k]$ corresponding to the factor system

$$\omega^{G_0}_{M,k}(g',g)=\exp(iM(\tfrac{1}{2}b\mathbf{v}'^2+\mathbf{v}'\cdot\mathbf{a}^{\phi'}))\times\exp(-i(k/2)(\mathbf{v}'\wedge\mathbf{v}^{\phi'})_3). \tag{4.5}$$

The calculation of $H^2{}_{*G}(G,\mathrm{U}(1))$ is similar to the one in the previous example. The result depends on the subgroup H of G which is represented by linear operators. The computations are straightforward and we only state the results in Table I.

From Table I we see that in the physically relevant case, $\mathrm{U}(P)$ linear and $\mathrm{U}(T)$ antilinear, only factor systems with $k=0$ appear as restrictions to G_0 of factor systems of G. The representation of the line reflection P through a linear operator $\mathrm{U}(P)$ prevents the possibility of factor systems with $k\neq0$, which, as we know, are not to be found in LOR's of G_0.

We remark that the "interaction" part $\Lambda(g,\alpha')$ is always missed although the corresponding equations do not imply $\Lambda(g,'\alpha)=1$ but only mean that $\Lambda(g,\alpha')$ is a two-coboundary.

The structure of the representation group $\overline{G}$ is similar, but *not* identical to the one for the $(3+1)$ case. The reason for that difference is the known fact that in two spatial dimensions the point reflection is a rotation, unlike the three-dimensional case. The composition law of $\overline{G}$ is

$$(\theta,g',(\gamma',\alpha'))(\theta,g,(\gamma,\alpha)) = (\theta'+\theta^{\alpha'}+\omega^{G_0}_M(g',g^{\alpha}),g'g^{\alpha'},(\gamma',\alpha')(\gamma,\alpha)), \tag{4.6}$$

with $\theta^{\alpha}=\theta$ for $\alpha=1,P$, and $\theta^{\alpha}=-\theta$ for $\alpha=T,PT$, and (γ,α) denotes an element of the representation group $\overline{V}$ for V (see Ref. 16).

This group acts on the $(2+1)$ space-time through the projection $(\theta,b,\mathbf{a},\mathbf{v},\phi,\gamma,\alpha)\to(b,\mathbf{a},\mathbf{v},\phi,\alpha)$. The isotopy subgroup of the point $(0,0)$ is $\overline{\Gamma}=\{(\theta,0,\mathbf{0},\mathbf{v},\phi,\gamma,\alpha)\}$, and the subgroup $\overline{\Gamma}_+$ to be represented linearly is isomorphic to $\mathbf{R}\otimes(\mathrm{E}(2)\odot\overline{V}_+)$, where $\overline{V}_+$ is the subgroup $\{(\gamma,\alpha)\}$ of $\overline{V}$ for which $\alpha=1$ or P. This subgroup $\overline{V}_+$ is isomorphic to $\mathbf{Z}_4\otimes\mathbf{Z}_2$ and acts on E(2) via the projection $\overline{V}_+\to V_+$ [i.e., (γ,α) acts as P and $(\gamma,1)$ as the identity]. In the isomorphism $\overline{V}_+\cong\mathbf{Z}_4\otimes\mathbf{Z}_2$, the first subgroup is generated by $(1,P)$ [$(1,P)^2=(\mu\nu,1)$ in the notation used in Ref. 13], and $\mathbf{Z}_2$ is the remaining part of V_+.

The linear one-dimensional representations of $\overline{\Gamma}_+$ are easily found. They are given by $U(\theta,\mathbf{v},\phi,\gamma,\alpha)=e^{iM\theta}\Delta_{\epsilon_1\epsilon_2\epsilon}(\gamma,\alpha)$ (same notations as in Ref. 13) and from them one gets the one-dimensional, and some two-dimensional semilinear representations of $\overline{\Gamma}$. With the choice $g_0=(0,\mathbf{0},0,0,1,T)\in\overline{\Gamma}-\overline{\Gamma}_+$ and proceeding as usual, we find that only those one-dimensional representations of $\overline{\Gamma}_+$ with $\epsilon_1=\epsilon_2=1$ are Wigner type I. Hence one-dimensional semilinear representations of $\overline{\Gamma}$ are parametrized by (M,ϵ), with $M\in\mathbf{R}$ and $\epsilon=\pm1$, and they are given by

$$\Delta_{\epsilon M}(\theta,\mathbf{v},\phi,\gamma,1)=e^{iM\theta};\quad \Delta_{\epsilon M}(0,\mathbf{0},0,1,P)=\epsilon;\quad \Delta_{\epsilon M}(0,\mathbf{0},0,1,T)=\mathbf{K}, \tag{4.7}$$

$\mathbf{K}$ being the complex conjugation operator.

Proceeding in the same way, one easily finds out the corresponding represenations of $\overline{G}$. They are parametrized by $[\beta,\epsilon]$, $\beta\in\mathbf{R}$, $\epsilon\in\{1,-1\}$ and its explicit expression is

$$D_{\beta\epsilon}(\theta,b,\mathbf{a},\mathbf{v},\phi,\gamma,1)=e^{i\beta b};\quad D_{\beta\epsilon}(0,0,\mathbf{0},\mathbf{0},0,1,P)=\epsilon,\quad D_{\beta\epsilon}(0,0,\mathbf{0},\mathbf{0},0,1,T)=\mathbf{K} \tag{4.8}$$

so that we obtain that the classes of LOR of the $(2+1)$ complete Galilei group with P linearly and T antilinearly represented are given by induction from semilinear representations of Γ

$$(U(g,\alpha)f)((g,\alpha)\mathbf{x})=\exp(iM(\tfrac{1}{2}b\mathbf{v}'^2+\mathbf{v}'\cdot\mathbf{a}^{\phi^{\alpha}}))\Lambda(\alpha)f(\mathbf{x}), \tag{4.9}$$

with Λ being a semilinear representation of $\overline{V}$. Here we can choose the section $\bar{s}$ as $\bar{s}(x)=(0,\mathbf{x},\mathbf{0},0,1,1)$.

As pointed out at the beginning of the example, things

TABLE I. Factor systems for Galilei $(2+1)$ group.

$\mathrm{U}(P)$	$\mathrm{U}(T)$	General factor system of G equivalent to	$H^2_{*G}(G,\mathrm{U}(1))$
unitary	antiunitary	$\omega^{G_0}_{M,0}(g',g^{\alpha})\omega^{V}_{m,n}(\alpha',\alpha)$	$\mathbf{R}\otimes\mathbf{Z}_2\otimes\mathbf{Z}_2$
antiunitary	unitary	$\omega^{G_0}_{0,k}(g',g^{\alpha})\omega^{V}_{m,n}(\alpha',\alpha)$	$\mathbf{R}\otimes\mathbf{Z}_2\otimes\mathbf{Z}_2$
antiunitary	antiunitary	$\omega^{V}_{m,n}(\alpha',\alpha)$	$\mathbf{Z}_2\otimes\mathbf{Z}_2$

are very different if one searchs for LOR with P antilinear, for then necessarily $M = 0$.

2. Newton–Hooke 3+1 groups

The connected Hooke group H_0 has been studied in Ref. 9. Here we will study the LOR of the complete group taking into due account the reflections. This study is also relevant to the "Dirac-like" equation for this case, the so-called Dubois equation.[9,17]

We will only consider the oscillating group H^-, hereafter referred to simply as H; the discussion for H^+ is very similar. As in the preceding examples we are going to consider only the realizations of H which represent linearly the subgroup $H_+ = H_0 \odot V_+$, with $V_+ = \{1,P\}$. The usual way of computing the group $H^2{}_{*H}(H,\mathrm{U}(1))$ gives $H^2{}_{*H}(H,\mathrm{U}(1)) \cong \mathbf{R} \otimes \mathbf{Z}_2 \otimes V$, and a lifting of $[M,l,m,n]$, $M \in \mathbf{R}$, $l,m,n, \in \{1,-1\}$, has a mass M part $\omega^{H_0}{}_M(g,g^{\alpha'})$ (see Ref. 18 for an explicit expression of $\omega^{H_0}{}_M$) and the remaining terms are just as in the Galilean case.

The representation group is written in full[19] [where $g = (b,\mathbf{a},\mathbf{v},A)$, $A \in \mathrm{SU}(2)$]

$$\begin{aligned}(\theta',g',\gamma',\alpha')(\theta,g,\gamma,\alpha) &= \Big(\theta' + \theta^{\alpha'} + \frac{1}{2}\Big(\mathbf{v}'^2 - \frac{1}{\tau^2}\mathbf{a}'^2\Big)\tau \cos\frac{b^{\alpha'}}{\tau}\sin\frac{b^{\alpha'}}{\tau} \\ &\quad + A'\mathbf{a}^{\alpha'}\Big(\mathbf{v}'\cos\frac{b^{\alpha'}}{\tau} - \frac{\mathbf{a}'}{\tau}\sin\frac{b^{\alpha'}}{\tau}\Big) \\ &\quad - \mathbf{v}'\mathbf{a}'\sin^2\frac{b^{\alpha'}}{\tau}, g'g^{\alpha'}, (\gamma',\alpha')(\gamma,a)\Big),\end{aligned} \tag{4.10}$$

where θ transforms under $\overline{V}$ via the projection $\overline{V} \to \overline{V}_+$ "as a time." The isotopy group $\overline{\Gamma}$ of the point $(0,\mathbf{0})$ is the set $\{(\theta,0,\mathbf{0},\mathbf{v},A,\gamma,\alpha)\}$ whose group structure is $\{\mathbf{R} \otimes (\mathbf{R}^3 \odot \mathrm{SU}(2)\} \odot V$, the subgroup $\mathbf{R}^3 \odot \mathrm{SU}(2)$ being isomorphic to the universal covering of the Euclidean three-dimensional group generated by pure Hooke transformations and rotations. The subgroup $\overline{\Gamma}_+$ is obtained by restriction of $\overline{V}$ to $\overline{V}_+$.

The study of one-dimensional representations of $\overline{\Gamma}$ and $\overline{H}$ is easily carried out, and the result is that the classes of one-dimensional semilinear representations of $\overline{\Gamma}$ modulo those that can be extended to $\overline{H}$ are parametrized by $M \in \mathbf{R}$, and a repersentative in each class is, for example, $\lambda(\theta,g,\gamma,\alpha) = e^{iM\theta}$. In addition to M, each LOR of $\overline{H}$ is specified by a pseudoequivalence class of finite-dimensional semilinear representations of $\overline{\Gamma}$. A particularly important class of LOR of H arises from four-dimensional representations of $\overline{\Gamma}$ [we note that four is the minimal dimension in order to get faithful representations of $\mathbf{R}^3 \odot \mathrm{SU}(2)$]. We choose the following representation of $\overline{\Gamma}_+$:

$$\sigma_{M(1/2)\epsilon_1\epsilon_2\epsilon}(\theta,\mathbf{v},A,\gamma,\alpha) = e^{iM\theta} D_{1/2}(\overline{v},A)\Delta_{\epsilon_1\epsilon_2\epsilon}(\gamma,\alpha), \tag{4.11}$$

where

$$D_{1/2}(\mathbf{v},A) = \begin{pmatrix} D_{1/2}(A) & 0 \\ \frac{1}{2}\boldsymbol{\sigma}\cdot\mathbf{v}D_{1/2}(A) & D_{1/2}(A) \end{pmatrix}, \tag{4.12}$$

with $\boldsymbol{\sigma} \equiv (\sigma_1,\sigma_2,\sigma_3)$ the Pauli matrices, and for the $\overline{V}_+$ subgroup generated by $(\mu,1)$, $(\nu,1)$, and $(1,P)$,

$$\Delta_{\epsilon_1\epsilon_2\epsilon}(\mu,1) = \epsilon_1; \quad \Delta_{\epsilon_1\epsilon_2\epsilon}(\nu,1) = \epsilon_1;$$

$$\Delta_{\epsilon_1\epsilon_2\epsilon}(1,P) = \begin{cases} \begin{pmatrix} \epsilon\mathbf{1}_2 & \\ & -\epsilon\mathbf{1}_2 \end{pmatrix} & \text{if } \epsilon_1\epsilon_2 = 1, \\ \begin{pmatrix} i\epsilon\mathbf{1}_2 & \\ & -i\epsilon\mathbf{1} \end{pmatrix} & \text{if } \epsilon_1\epsilon_2 = -1. \end{cases} \tag{4.13}$$

Of course when restricted to the connected group this reduces, no matter what the values of $\epsilon_1,\epsilon_2,\epsilon$, to the ordinary $[M,1/2]$ representation of $\mathbf{R} \otimes (\mathbf{R}^3 \odot \mathrm{SU}(2))$ which is known to be linked with particles of mass M and spin $\frac{1}{2}$ because of the presence of a factor system for mass M and of a restriction to SU(2) which is a multiple of $D_{1/2}$.

The next step is to induce a semilinear representation of $\overline{\Gamma}$. Let us choose $g_0 = (0,\mathbf{0},\mathbf{1},1,PT)$; the automorphism $\overline{\Gamma}_+ \to \overline{\Gamma}_+$ induced by the inner conjugation in $\overline{\Gamma}$ corresponding to g_0 is simply given by

$$(\theta,\mathbf{v},A,\gamma,\alpha) \to (-\theta,\mathbf{v},A,\gamma,\alpha).$$

TABLE II. Representations $\sigma_{M(1/2)\epsilon_1\epsilon_2}$ of the group Γ up to pseudoequivalence.

Indices of pseudoequivalence classes ϵ_1	ϵ_2	Dimension	$(0,\mathbf{v},A,1,1)$	$(0,\mathbf{0},\mathbf{1},1,P)$	$(0,\mathbf{0},\mathbf{1},1,PT)$
-1	-1	4	$D_{1/2}(\mathbf{v},A)$,	$\begin{pmatrix} \mathbf{1}_2 & \\ & -\mathbf{1}_2 \end{pmatrix}$,	$\begin{pmatrix} \tau & \\ & -\tau \end{pmatrix}\mathbf{K}$
1	1	8	$\begin{pmatrix} D_{1/2}(\mathbf{v},A) & \\ & D_{1/2}(\mathbf{v},A) \end{pmatrix}$,	$\begin{pmatrix} \mathbf{1}_2 & & & \\ & -\mathbf{1}_2 & & \\ & & \mathbf{1}_2 & \\ & & & -\mathbf{1}_2 \end{pmatrix}$,	$\begin{pmatrix} & & \tau & \\ & & & -\tau \\ \tau & & & \\ & -\tau & & \end{pmatrix}$
1	-1	8	$\begin{pmatrix} D_{1/2}(\mathbf{v},A) & \\ & D^*_{1/2}(\mathbf{v},A) \end{pmatrix}$,	$i\begin{pmatrix} \mathbf{1}_2 & & & \\ & -\mathbf{1}_2 & & \\ & & \mathbf{1}_2 & \\ & & & -\mathbf{1}_2 \end{pmatrix}$,	$\begin{pmatrix} & \mathbf{1}_4 \\ \mathbf{1}_4 & \end{pmatrix}\mathbf{K}$
-1	1	8	$\begin{pmatrix} D_{1/2}(\mathbf{v},A) & \\ & D^*_{1/2}(\mathbf{v},A) \end{pmatrix}$,	$i\begin{pmatrix} \mathbf{1}_2 & & & \\ & -\mathbf{1}_2 & & \\ & & \mathbf{1}_2 & \\ & & & -\mathbf{1}_2 \end{pmatrix}$	$\begin{pmatrix} & -\mathbf{1}_4 \\ \mathbf{1}_4 & \end{pmatrix}\mathbf{K}$

with $\tau = \begin{pmatrix} & 1 \\ -1 & \end{pmatrix} = i\sigma_y$.

Hence the g_0 transformed of the representation $\{M,1/2,\epsilon_1,\epsilon_2,\epsilon\}$ is $\{-M,1/2,\epsilon_1,\epsilon_2,\epsilon\}$ and we have to search for its antiunitary equivalence/inequivalence. Some calculations give the results displayed in Table II. In fact when $\epsilon_1\epsilon_2 = 1$ both are antiunitarily equivalent by means of the complex conjugation $\mathbf{K}$, and their Wigner type is I for $\epsilon_1 = -1$ and II for $\epsilon_1 = 1$. When $\epsilon_1\epsilon_2 = -1$ they are not equivalent.

Furthermore, representations with different M and/or ϵ are obviously pseudoequivalent, so that both indices disappear from the labeling for pseudoequivalence classes. For $\Delta(T)^2 = \Delta(PT)^2 = -1$ the representation remains four-dimensional. In all other cases the dimension is doubled from four to eight, and the representation is fully characterized (as long as pseudoequivalence classes are concerned) by its restriction to $\mathbf{R}^3 \odot SU(2)$ and the values of the squares of the operators $\Delta(T)$ and $\Delta(PT)$. Notice that $\Delta(T) = \epsilon_1\epsilon_2\Delta(P)\Delta(PT)$.

From these expressions the explicit form of the corresponding LOR's of H are found from formula (3.4) or (3.5).

C. Invariance groups of electromagnetic fields

These groups have been studied from several viewpoints,[1,2,20–24] both for the relativistic and nonrelativistic case.

As an example we consider the symmetry group of a uniform electric field, along the z axis, in the nonrelativistic case. The connected symmetry group is the subgroup of Galilei $3+1$ generated by H, P_i, K_i, J_3, $i = 1,2,3$. Furthermore, time reversal is also a symmetry so that we can consider the group with time reversal. This is a semidirect product, $\mathbf{Z}_2 = \{1,T\}$ acting on the connected group as

$$T: (b,\mathbf{a},a_3,\mathbf{v},v_3,\phi) \to (-b,\mathbf{a},a_3,-\mathbf{v},-v_3,\phi),$$

where $\mathbf{a}$ denotes (a_1,a_2). If T is to be represented antilinearly, one obtains the group $H^2_{\mathrm{loc}*G}(G,\mathrm{U}(1))$ as $\mathbf{R}^3 \otimes \mathbf{Z}_2$; the $\mathbf{R}^3$ part comes from the connected case [for then $H^2_{\mathrm{loc}}(G_0,\mathrm{U}(1)) \cong \mathbf{R}^3$] and $\mathbf{Z}_2$ comes from the subgroup $\{1,T\}$. The local representation group is $\{(\theta,\zeta,\eta;b,\mathbf{a},a_3,\mathbf{v},v_3,\phi;\gamma,\alpha)\}$ where $\theta,\zeta,\eta \in \mathbf{R}$ and $\{\gamma,\alpha\}$ is the representation group $\overline{\mathbf{Z}}_2$ for $(\mathbf{Z}_2,\mathbf{Z}_1)$. The product law involves the explicit form of the factor systems and will not be written here. The complete discussion can be easily carried out, and the results include, of course, Kramers theorem [if $\mathrm{U}(T)^2 = -1$ then there cannot be nondegenerate levels for a particle in a homogeneous electric field.]

D. One-dimensional conformal group

Our last example is a one-dimensional conformal group.[25] This name ordinarily applies to the group of transformations of the (compactified) real line

$$\exp(bH): t \to t+b, \quad \exp(dD): t \to e^{-d}t,$$
$$\exp(\theta K): t \to t/(1-\theta t).$$

The Lie brackets are

$$[D,H] = -H, [D,K] = K, [H,K] = 2D.$$

In fact, this is the connected component of the identity in the group G obtained when one adjoins the simple inversion I: $t \to 1/t$ to the translation subgroup $\{\exp(bH), b \in \mathbf{R}\}$ [notice $\exp(\theta K) = I\exp(-\theta H)I$]. Alternatively we can enlarge the connected component of the identity by means of the point-reflection P: $t \to -t$; in both cases the group G obtained is the same. The group G [isomorphic to $GL(2,\mathbf{R})/\mathbf{R}^*$] is also a semidirect product of the connected conformal group [isomorphic as one knows to $SL(2,\mathbf{R})/\mathbf{Z}_2$] by a $\mathbf{Z}_2$ group which can be taken either $\{1,I\}$ or $\{1,P\}$. Some expressions are simpler when the last subgroup is chosen, so that we will use it. In terms of the previous parametrization, $\mathbf{Z}_2 = \{1,P\}$ acts as

$$P: (b,d,\theta) \to (-b,d,-\theta).$$

We shall use the standard notation (g,α) with $\alpha \in \{1,P\}$.

Let us now study those LOR's with P represented by means of an unitary operator. The group $H^2{}_{*G}(G,\mathrm{U}(1))$ is trivial, because the corresponding group for the connected component of identity is already known to be trivial, the part for the subgroup $\{1,P\}$ is trivial, too, and the interaction part $\Lambda(g,\alpha)$ must be, for fixed α, a one-dimensional representation of the connected component of the identity which is also necessarily trivial.

Hence the group G is its own representation group. The isotopy group is the subgroup $\{0,d,\theta,\alpha\}$ whose composition law is

$$(d',\theta',\alpha)(d,\theta,\alpha) = (d',\theta')(d,\theta)^{\alpha'},\alpha'\alpha),$$

with $(d,\theta)^\alpha = (d,\theta^\alpha)$, $\theta^P = -\theta$. Its one-dimensional representations are easily shown to be labeled by $r \in \mathbf{R}$, $\epsilon \in \{1,-1\}$, and given by

$$(d,\theta,\alpha) \to e^{idr}\Delta_\epsilon(\alpha) \text{ such that } \Delta_\epsilon(P) = \epsilon.$$

The crossed homomorphisms of G are trivial.

Finally, to obtain the LOR's of G is easy using formula (3.4). In this case the realizations of G are equivalent to representations.

ACKNOWLEDGMENTS

One of us (M.S.) thanks the Centre de Recherches Mathématiques de l'Université de Montréal for its hospitality and particularly Professor P. Winterniz.

The research reported in this paper was supported in part by the Comisión Asesora de Investigación Científica y Técnica and the Ministerio de Educación Ciencia (Grant to M. A. de Olmo from the Plan de Formación de Personal Investigador).

APPENDIX: FACTOR SYSTEMS OF SEMIDIRECT PRODUCTS

Theorem: Let G be a Lie group which is a semidirect product $G = G_0 \odot V$ with the action α: $g \to g^\alpha$, $g \in G_0$, $\alpha \in V$, and let H be a closed subgroup of G of index 1 or 2. The action of G on U(1) is denoted by $*H$, and their restrictions to G_0 and V give the actions of G_0 and V on U(1) denoted by $*(G_0 \cap H)$ and $*(V \cap H)$, respectively. Then each element $[\omega] \in H^2{}_{*H}(G,\mathrm{U}(1))$ has a lifting $\omega^G \in \mathbf{Z}^2{}_{*H}(G,\mathrm{U}(1))$ given by

$$\omega^G(g,\alpha;h,\beta) = \omega^{G_0}(g,h^\alpha)\,[\omega^V(\alpha,\beta)]^{*(H\cap G_0)(gh^\alpha)}\Lambda(g,\alpha)^{*(H\cap G_0)(g)}, \tag{A1}$$

where $\omega^{G_0}\in Z^2{}_{*(H\cap G)}(G_0U(1))$, $\omega^V\in Z^2{}_{*(H\cap V)}(V,U(1))$, and Λ: $G_0\times V\to U(1)$ is a Borel function verifying the following relationships:

$$\omega^{G_0}(g^\alpha,h^\alpha) = (\omega^{G_0}(g,h))^{*(H\cap V)(\alpha)}\Lambda(gh,\alpha) \times\{\Lambda(g,\alpha)[\Lambda(h\alpha)]^{*(H\cap G_0)(gh,\alpha)}\}^{-1}, \tag{A2}$$

$$\Lambda(g,\alpha\beta) = \Lambda(g\beta,\alpha)[\Lambda(g,\beta)]^{*(H\cap V)(\alpha)} \times[\omega^V(\alpha,\beta)]^{*(H\cap G_0)(g^{\alpha\beta})}[\omega^V(\alpha,\beta)]^{-1}. \tag{A3}$$

Conversely, let us take two actions of G_0 and V on U(1) and let H be the subgroup of index 1 or 2 generated by their kernels. If ω^{G_0}, w^V, and Λ are functions satifying the above relationships (A2) and (A3) then ω_G defined by (A1) lives in $Z^2{}_{*H}(G,U(1))$. The proof can be found in Ref. 25. (Other references are 26 and 27.)

A particular case of this theorem appears when G is a nonconnected Lie group whose structure is that of a semidirect product of its connected identity component G_0 by its group of connected components $V\cong\pi_0(G)$. Then the action $*(H\cap G_0)$ of G_0 in U(1) is trivial. The above relationship becomes

$$\omega^G(g,\alpha;h,\beta) = w^{G_0}(g,h^\alpha)\omega^V(\alpha,\beta)\Lambda(h,\alpha) \tag{A4}$$

with Λ: $G_0\times V\to U(1)$ verifying

$$\omega^{G_0}(g^\alpha,h^\alpha) = \omega^{G_0}(g,h)^{*(H\cap V)((\alpha)}\Lambda(gh,\alpha) \times\{\Lambda(g,\alpha)\Lambda(h,\alpha)\}^{-1}, \tag{A5}$$

$$\Lambda(g,\alpha\beta) = \Lambda(g^\beta,\alpha)[\Lambda(g,\beta)]^{*(H\cap V)(\alpha)}. \tag{A6}$$

[1]H. Hoogland, Nuovo Cimento B **32**, 427 (1976); J. Phys. A: Math. Gen. **11**, 797, 1557 (1978).
[2]H. Hoogland, Ph.D. thesis, University of Nijmegen, 1977.
[3]V. Bargman, Ann. Math. **59**, 1 (1954).
[4]E. P. Wigner, Ann. Math. **40**, 149 (1939).
[5]J. F. Cariñena, M. A. del Olmo, and M. Santander, Physica **114**, 420 (1982).
[6]M. Asorey, J. F. Cariñena, and M. A. del Olmo, J. Phys. A: Math. Gen. **16**, 1603 (1983).
[7]J. F. Cariñena, M. A. del Olmo, and M. Santander, *Lecture Notes in Physics*, Vol. 201, edited by G. Denardo (Springer, Berlin, 1983).
[8]J. F. Cariñena, M. A. del Olmo, and M. Santander, J. Phys. A: Math. Gen. **17**, 3091 (1984).
[9]J. F. Cariñena, M. A. del Olmo, and M. Santander, J. Math. Phys. **26**, 2096 (1985).
[10]G. W. Mackey, Ann. Math. **55**, 701 (1952) [see also D. J. Simms, "Lie groups and quantum mechanics," *Lecture Notes in Physics*, Vol. 52 (Springer, Berlin, 1969)].
[11]E. P. Wigner, "Unitary representations of the inhomogeneous Lorentz group including reflections," in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gursey (Gordon and Breach, New York, 1964).
[12]U. Cattaneo, J. Math. Phys. **19**, 452 (1978).
[13]J. F. Cariñena, and M. Santander, **20**, 2168 (1979).
[14]R. Shaw and J. Lever, Commun. Math. Phys. **38**, 257 (1974).
[15]J. M. Levy-Leblond, "Galilei Group and Galilei invariance," in *Group Theory and its Applications*, edited by E. Loebl (Academic New York, 1971), Vol. II.
[16]J. F. Cariñena, and M. Santander, J. Math. Phys. **22**, 1548 (1981).
[17]J. G. Dubois, Can. J. Phys. **51**, 1757 (1973).
[18]J. R. Derome and J. G. Dubois, Nuovo Cimento B **9**, 351 (1972).
[19]M. Valls, Tesina de Licenciatura, Universidad de Valladolid, 1977 (unpublished).
[20]H. Bacry, Ph. Combe, and J. L. Richard, Nuovo Cimento A **67**, 267 (1970); **70**, 289 (1970).
[21]Ph. Combe and J. L. Richard, *Proceedings of the 2nd International Colloquium on Group Theoretical Methods in Physics*, Vol. 2, edited by A. Janner and T. Jannsen (University of Nijmegen, Nijmegen, 1973).
[22]N. Giovannini, Helv. Phys. Acta **50**, 337, 349 (1977); Physica A **87**, 546 (1977).
[23]V. Hussin, Nuovo Cimento A **65**, 39 (1981).
[24]J. Beckers, and V. Hussin, J. Math. Phys. **24**, 1286, 1295 (1983).
[25]K. R. Parthasaraty, *Multiplier on Locally Compact Groups*, Lecture Notes in Physics, Vol. 93 (Springer, New York, 1969).
[26]P. M. van den Broeck, J. Phys. A: Math. Gen. **9**, 855 (1976).
[27]M. Santander, Ph.D. thesis, University of Valladolid, 1977 (unpublished).

Substitutional symmetries and extra degeneracies of real symmetric matrices

T. H. Siddall
Chemistry Department, University of New Orleans, New Orleans, Louisiana 70148
J. J. Sullivan
Physics Department, University of New Orleans, New Orleans, Louisiana 70148

(Received 9 December 1986; accepted for publication 20 January 1988)

The relation between substitutional symmetry operations that leave a real symmetric matrix invariant and the degeneracies exhibited by the matrix in diagonal form are examined. The usual application of group theory to this problem is formulated. Substitutional (and other) symmetries can exist, which do not form part of the invariance group. These symmetries can cause extra degeneracy of the root system and are frequently encountered in some physical applications such as the Hückel model of molecular bonds. Some general features that lead to extra degeneracy are noted and illustrative examples are given for systems of six equivalent centers.

I. INTRODUCTION

Group theory is extensively used to block diagonalize matrices possessing a group symmetry and thus enumerate and specify the degeneracy of the root system. Reduction of a carrier space to its irreducible representations (irreps) by the use of projection operators determines not only the degeneracies but the roots and their eigenvectors up to equivalence within the resolution of the multiplicity of a given irrep. Degenerate eigenvectors are distinguished by choosing diagonalization with respect to a specified subgroup chain. Chemists and physicists are so used to the idea that knowledge of the symmetry group implies knowledge of the degeneracies that the existence of unexplained systematic degeneracies in a root system usually stimulates a search for hidden symmetries and higher groups. The application of $O(4)$ to the bound states of the Coulomb problem[1] and the unitary group to the harmonic oscillator[2] are well known examples of the success of such efforts. It is also well known that accidental degeneracies occur when particular values of model parameters force crossings of energy levels for different symmetry species. Often correlation tables and some knowledge of the energy level spacings in two high symmetry limits are sufficient to show that accidental degeneracies must occur in a model. While crossings in parameter space can be more general, accidental degeneracies usually require particular numerical sets of values for the model parameters. We use the term systematic degeneracy to signify degeneracy that persists over a continuous range of parameter values. In this paper we examine the question "Do systematic degeneracies necessarily imply a group structure?" Counterexamples derived for specific models have already been reported in the literature showing that the question must be answered in the negative.[3]

More specifically we examine the root structure of real symmetric $n \times n$ matrices with substitutional symmetries among the elements. Our attention was drawn to this problem in terms of molecular orbital calculations using a simple Hückel model. Using this model it is common to obtain a root structure with more degeneracy than can be implied by the spatial point group symmetry.[4] Similar matrices and higher degeneracies than those derivable from simple point group analysis frequently arise in vibrational or electron orbital calculations when models are subjected to various restrictions.

We note at the start that the root system of a matrix remains the same under all similarity transformations. A matrix in its equivalent forms will in general display different substitutional symmetry. Therefore it is not surprising that different substitutional symmetries can be connected to the same degeneracy system, i.e., even if systematic degeneracies implied a substitutional group structure it would not necessarily be unique. However, one might hope that knowledge of the eigenvectors would be sufficient to distinguish between possible symmetry groups and one could determine the matrix form with maximum substitutional symmetry. In the next section we establish a general framework within which the substitutional symmetry of real symmetric matrices may be analyzed. By specific construction we show that two distinct substitutional symmetry groups can result in the same degeneracies. By identification of the roots these symmetry groups can apply to equivalent forms of the same matrix. By specific construction we also show how higher degeneracies can exist without necessarily implying a higher symmetry group. Section III presents illustrative examples using 6×6 matrices. We close with some remarks.

II. GENERAL THEORY

We wish to establish the degeneracies and symmetry species of the root form of an $n \times n$ real symmetric matrix which is invariant under a set of substitutional symmetry operations. For some physical applications one can easily specialize to matrices with zero diagonal elements. Such matrices are encountered in molecular vibration theory and molecular orbital theory as in the Hückel approximation. Since we are considering only substitutional symmetries, we may limit analysis to permutational groups; i.e., S_n, the symmetric group on n items, and its subgroups. All three-dimensional point groups are isomorphic to subgroups of the sym-

 0022-2488/88/051091-06$02.50

metric group, but the latter has a much richer group structure. Indeed the symmetric group S_n is isomorphic to an $(n-1)$-dimensional point group as can be shown by an inductive construction. Thus S_2 describes two equivalent points of a line that generates the planar equilateral triangle with S_3 symmetry which in turn generates the tetrahedral symmetry S_4, etc. In general the matrix may have substitutional symmetry above the three-dimensional space symmetry of the model. The full symmetry group may have no point group realization although it does contain the geometric symmetry as a subgroup. Even considering these higher permutational groups we will show that some systematic degeneracies are due to symmetries that cannot be represented as elements of a group.

The natural representation on n items has $n \times n$ permutational matrices with zero entries except for a single unit in any row or column. If a permutation takes j to k, its matrix representation has a unit in the k th row of the j th column. In reduced form in S_n the natural representation spans $[n]+[n-1,1]$, the one-dimensional totally symmetric irrep and the $(n-1)$- dimensional defining irrep. The matrices of the natural representation span the Kronecker square of the representation which further separates into symmetrized and antisymmetrized forms. Symmetric matrices span the irreps $([n]+[n-1,1])\odot[2]=2[n]+2[n-1,1]+[n-2,2]$. The diagonal elements themselves transform as the natural representation so real symmetric matrices with zeros on the diagonal transform as $[n]+[n-1,1]+[n-2,2]$. The above decomposition requires that a real symmetric matrix invariant under all the permutations of S_n can be considered as a linear combination of at most two invariant forms (corresponding to the multiplicity of the invariant species $[n]$ in the symmetrized square) and has roots that form a singlet and an $n-1$ degenerate level (corresponding to the decomposition of the natural representation). Requiring the additional condition of zero values on the diagonal restricts the number of invariant matrices in this case to one. The group theoretical classification of the number of invariant forms and the resultant degeneracy of the roots proceeds in the same way for any subgroup. The irrep decomposition of the natural representation on restriction to the subgroup determines the group caused degeneracy of the roots. The number of possible invariants equal the multiplicity of the totally symmetric irrep in the decomposition of the Kronecker square (or equivalently in the restriction of $2[n]+2[n-1,1]+[n-2,2]$ to the subgroup). The number of diagonal invariants is simply the multiplicity of the totally symmetric irrep in the decomposition of the natural representation so the number of invariants with zeros on the diagonal is easily found. Because we deal only with real matrices only real irreps or complex representations in pairs will occur in the decomposition of the natural representation. Therefore the multiplicity of the totally symmetric irrep in the symmetrized Kronecker square is equal to the sum of binomials $\binom{f+1}{2}$, where f is the multiplicity of an irrep in the decomposition of the natural representation. For zeros on the diagonal one need only sum the binomials $\binom{f-1}{2}$.

Group theory allows more than just enumeration of the number of invariants and the resulting degeneracies. To uniquely specify the eigenvectors of the root form one simultaneously diagonalizes a maximum set of mutually commuting but independent operators. Within this set are invariant operators corresponding to the class sums, which, when diagonalized, serve to specify the irreps. Subclass operators invariant within a subgroup sequence serve to specify the basis within a degenerate species. The invariant matrices representing class sums within the natural representation are easily determined. Character theory requires their matrix within an irrep be a constant times the identity matrix. The constant equals the class order times the irrep character of the class divided by the irrep dimension, that is, the invariant matrices corresponding to class sums in a group or its subgroups are known in the natural permutation representation and in their irrep root forms. There are as many independent invariant matrices representing class sums as there are inequivalent irreps occurring in the decomposition of the natural representation. Multiplicities usually cannot be resolved solely by group theory and generally require solving for the roots of a polynomial the order of the multiplicity. As an elementary example there are only two independent matrices representing class sums in the natural representation of S_n. These can be taken as the identity matrix E_n and a matrix with the unit in all off-diagonal positions and zeros on the diagonal. Character theory requires that this latter matrix has root form of a singlet with value $n-1$ and an $n-1$ degenerate level with root -1. In S_n our problem is completely determined by group theory. Group theory also completely determines the diagonalization when the subgroup involved is the cyclic group C_n. This is because the natural representation of S_n spans the regular representation of C_n, which has only one-dimensional (Fourier) irreps and therefore no multiplicity. Actually the point group frequently encountered is C_{nv} of a planar ring structure or its equivalent D_{md} $(n=2m)$ for a puckered ring. Either of these point groups are isomorphic to the semidirect product $C_n V S_2$. The effect of the semidirect product is to join the cycle element $(n)^p$ and its inverse $(n)^{n-p}$, $0<p<n/2$, into the same class (p), and to join the Fourier irrep k with its complex conjugate irrep $n-k$, $0<k<n/2$, into the two- dimensional real irrep $\{k\}$. There is still no multiplicity in the decomposition of the natural representation and group theory gives the roots $\{k\}=a_0+a_p(\omega^{kp}+\omega^{-kp})$, where $\omega=\exp(i2\pi/n)$ and a_p is the pth neighbor interaction. For $n/2=m=p$ the quantity in parentheses should be set equal to $(-1)^k$. The invariant matrix has elements $A_{ij}=a_{|i-j|}$. The expression for the roots is easily understood in terms of the class sums with the matrix representing the cycle (n) in diagonal form. Group theory and a knowledge of the class invariant matrices completely solves the multiplicity-free problem.

The application of group theory to this problem may be summarized in the following steps.

(1) Decompose the natural representation of S_n on restriction to the spatial subgroup $([n]+[n-1,1])\downarrow \simeq \Sigma[\lambda_i]f_i$. The dimensions of the irreps give the degeneracies of the roots. The roots of irreps occurring without multiplicity are completely determined by the class matrices as simple linear combinations of the invariant matrix param-

eters. The roots of irreps occurring with multiplicity f must be determined from a polynomial (the secular determinant) of rank f in the invariant matrix parameters.

(2) The number of invariant matrix parameters is equal to the sum of the binomials $\binom{f+1}{2}$.

It is instructive to examine the diagonalization procedure in a little more detail. The rows of the transformation matrix that bring about the block diagonal form within an equivalent irrep species are essentially proportional to the projection operators of the group. Indeed the projection operators defined for a general group G as

$$|\lambda mn\rangle\langle g| = \{|\lambda|/|G|\}^{1/2}\Sigma\begin{bmatrix}\lambda & g\\ m & n\end{bmatrix}^*\langle g| \quad (2.1)$$

can be considered as the transformation matrix reducing the regular representation of the group. For our case only the projection operators for the irreps in the reduction of the natural representation will occur. The size of the resulting block matrices is the multiplicity of the irrep species in the reduction with identical blocks appearing equal to the dimension of the irrep. Note the following similarity transformation for submatrices of an overall matrix:

$$\begin{bmatrix}U_1 & 0\\ 0 & U_2\end{bmatrix}\begin{bmatrix}A & C\\ C^+ & B\end{bmatrix}\begin{bmatrix}U_1^+ & 0\\ 0 & U_2^+\end{bmatrix} = \begin{bmatrix}U_1AU_1^+ & U_1CU_2^+\\ U_2C^+U_1^+ & U_2BU_2^+\end{bmatrix}. \quad (2.2)$$

Under $S_{n-m}S_m$ the natural representation spans $(2[n-m]+[n-m-1,1])\ [m]+[n-m]\ [m-1,1]$ so for $m>1$ there are five invariants,two diagonals, two off-diagonals connecting elements within each subset, and one off-diagonal connecting elements of the two subsets. Let a and a' be the two diagonal parameters, b and b' be the off-diagonal parameters within the subsets, and c be the parameter between the subsets. Block diagonalization is obtained by using projection operators for the constituent irreps. Block diagonalization comes about because the $n-m\times m$ matrix representing the interaction between the subsets (with c in every position) transforms only as $[n-m]\ [m]$ and the other projections give zeros in these positions. Thus we find in the block diagonal form $a-b$ (1×1)'s occurring $n-m-1$ times, $a'-b'$ occurring $m-1$ times, and a 2×2 with entries

$$\begin{bmatrix}a+(n-m-1)b & \{(n-m)mc^2\}^{1/2}\\ \{(n-m)mc^2\}^{1/2} & a'+(m-1)b'\end{bmatrix}.$$

Solving the quadratic this has roots

$$\{a+(n-m-1)b+a'+(m-1)b'\}/2 \pm\{[(a+(n-m-1)b-a'-(m-1)b')/2]^2 + (n-m)mc^2\}^{1/2}. \quad (2.3)$$

We see with $a=a'$ and $b=b'=c$ this has the form of an S_n invariant and the corresponding root system discussed above. A special case occurs when $n=2m$ for which we may embed the symmetrized wreath product S_m wr S_2 in the subgroup chain. The natural representation carries the irreps $[m]\otimes([2]+[1^2])+([m],[m-1,1])$ giving three parameters with two singlets and an $(2m-2)$-fold degenerate level. The roots are completely determined by group theory. In particular no quadratic polynomial must be solved and no radical appears in the expression for the roots. Root expressions for this case are obtained from (2.3) by setting $a=a'$ and $b=b'$.

This same set of two singlets and a $2m-2$ degenerate level is obtained for the subgroup $S_{2m-1}S_1$, which, however, has four invariant parameters. Moreover, in this latter case a quadratic must be solved with the characteristic radical appearing in the expression for the two singlet roots. Thus we conclude the system of degeneracies does not uniquely determine the substitutional symmetry group of the matrix from which the roots were obtained. The expressions for the roots for the case $S_{n-1}S_1$ may be obtained from (2.3) by setting $m=1$ and $b'=0$. Since the three roots for $S_{2m-1}S_1$ are expressed in terms of four parameters, we can adjust the numerical values of these parameters so the roots are equal to those determined for an invariant matrix with S_m wr S_2 symmetry; i.e., the two matrices with different substitutional symmetries are in fact equivalent. Finally by returning to the general case for $S_{n-m}S_m$ and simply setting $a-b=a'-b'$ we obtain an $(n-2)$-fold degeneracy. The number of parameters is effectively reduced to four but there is no higher substitutional symmetry group for real symmetric matrices having this as its invariant form. Hence we must conclude the higher degeneracy does not result from purely group theoretical considerations.

III. EXAMPLE SYSTEMS

As particular examples we consider several subgroup sequences with $n=6$. The dimension is sufficient to make the examples nontrivial and illustrate all the points raised above. In addition several of the subgroups are isomorphic to point groups and therefore have direct physical application. We will make frequent reference to wreath product subgroups of the type S_n wr S_m. The wreath product group [a semidirect product $(S_n)^m \vee S_m$] is discussed extensively in the book by James and Kerber.[5] For spatial realizations and for ordering the matrices we will generally pair $(j,7-j)$. Table I gives the subgroup chains considered, the irrep decompositions, and analytic expressions for the roots as functions of the invariant matrix parameters and labelled by the irrep species. One may easily check the invariance of the matrix trace in the three examples given.

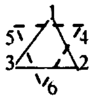
For the first example given in part (a) of Table I a spatial realization of S_3S_2 is given by two identical equilateral triangles in parallel planes with a sixty degree phase shift and a variable displacement along the perpendicular joining their midpoints (a trigonal antiprism). The lower group characterizes the substitutional symmetry of structures like the benzene ring. At a displacement such that the intervertex distances become identical the vertices are at the face centers of a cube with symmetry S_2 wr S_3. If on the other hand the two equilateral triangles rotate sufficiently fast about the perpendicular, the interactions between the vertices of the two triangles might be considered equivalent (isodynamic equivalences) and S_3 wr S_2 is appropriate although it has no point group realization. Setting $c=d$ the two doublets combine to form the quartet of S_3 wr S_2. Alternately setting $b=d$ the final doublet and singlet combine to form the triplet of S_2 wr S_3. With $b=c=d$ the two doublets and final

TABLE I. Three examples using subgroups of S_6. Presented in the order: possible subgroup chains; a diagram suggesting the spatial group; the subduction chains for the irrep [5, 1] of S_6; the form and parametrization of the invariant matrix, and analytic expressions for the roots of the invariant matrix.

(a) Subgroup D_{3d} including roots of the class matrices used to determine the irrep roots

Group chain

$$S_6 \nearrow S_3\,wr\,S_2 \searrow \quad S_6 \searrow S_2\,wr\,S_3 \simeq O_h \nearrow \quad S_3S_2 \simeq D_6 \simeq D_{3d} \simeq C_{6v}$$



Subduction chain

$$[5,1] \nearrow [3]\otimes[1^2] + ([3],[2,1]) \searrow \quad [5,1] \searrow [2]\otimes[2,1] + ([2]\otimes[2],[1^2]) \nearrow \quad [3][1^2] + [2,1]([2]+[1^2]).$$

Invariant matrix

$$\begin{bmatrix} a & b & b & d & d & c \\ b & a & b & d & c & d \\ b & b & a & c & d & d \\ d & d & c & a & b & b \\ d & c & d & b & a & b \\ c & d & d & b & b & a \end{bmatrix}$$

Irrep roots of the class matrices in the subgroup S_3S_2

class	parameter	[3][2]	[2,1][2]	[2,1][1^2]	[3][1^2]
$(1^6)=(1^3)(1^2)$	a	1	1	1	1
$(3^2)=(3)(1^2)$	b	2	-1	-1	2
$(2^3)=(1^3)(2)$	c	1	1	-1	-1
$(6)=(3)(2)$	d	2	-1	1	-2

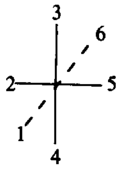
Irrep roots

$[3]\ [2] = a+2b+c+2d$ (singlet)
$[2,1]\ [2] = a-b+c-d$ (doublet)
$[2,1]\ [1^2] = a-b-c+d$ (doublet)
$[3]\ [1^2] = a+2b-c-2d$ (singlet)

(b) Subgroup D_{4h}

Group chain

$$S_6 \nearrow S_2\,wr\,S_3 \simeq O_h \searrow \quad S_6 \searrow S_4S_2 \simeq O_h \nearrow \quad (S_2\,wr\,S_2)S_2 \simeq D_{4h}$$



Subduction chain

$$[5,1] \nearrow [2]\otimes[2,1] + ([2]\otimes[2],[1^2]) \searrow \quad [5,1] \searrow [4]([2]+[1^2]) + [3,1][2] \nearrow \quad ([2]\otimes[2])([2]+[1^2]) + ([2]\otimes[1^2])[2] + ([2],[1^2])[2].$$

Invariant matrix

$$\begin{bmatrix} a & c & c & c & c & e \\ c & b & d & d & f & c \\ c & d & b & f & d & c \\ c & d & f & b & d & c \\ c & f & d & d & b & c \\ e & c & c & c & c & a \end{bmatrix}$$

TABLE I. (*Continued.*)

Irrep roots

$([2]\otimes[2])[2] = (a+e+b+f+2d)/2 \pm \{[(a+e-b-f-2d)/2]^2 + 8c^2\}^{1/2}$ (singlets)

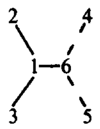
$([2]\otimes[1^2])[2] = b+f-2d$ (singlet)

$([2],[1^2])[2] = b-f$ (doublet)

$([2]\otimes[2])[1^2] = a-e$ (singlet)

(c) Subgroup D_{2d}

Group chain



$$\begin{tikzcd} & S_3 \text{ wr } S_2 \arrow[dr] & \\ S_6 \arrow[ur] \arrow[dr] & & (S_2S_1) \text{ wr } S_2 \simeq D_{2d} \\ & S_4S_2 \arrow[ur] & \end{tikzcd}$$

Subduction chain

$$\begin{tikzcd} & [3]\otimes[1^2] + ([3],[2,1]) \arrow[dr] & \\ {[5,1]} \arrow[ur] \arrow[dr] & & {[2]\otimes[2] + 2([2]\otimes[1^2]) + ([2],[1^2])} \\ & {[4]([2]+[1^2]) + [3,1][2]} \arrow[ur] & \end{tikzcd}$$

Invariant matrix

$$\begin{bmatrix} a & d & e & e & d & f \\ d & b & c & c & g & e \\ e & c & b & g & c & d \\ e & c & g & b & c & d \\ d & g & c & c & b & e \\ f & e & d & d & e & a \end{bmatrix}$$

Irrep roots

$[2]\otimes[2] = (b+g+2c+a+f)/2 \pm \{[(b+g+2c-a-f)/2]^2 + 2(d+e)^2\}^{1/2}$ (singlets)

$[2]\otimes[1^2] = (b+g-2c+a-f)/2 \pm \{[(b+g-2c-a+f)/2]^2 + 2(d-e)^2\}^{1/2}$ (singlets)

$([2],[1^2]) = b-a$ (doublets)

singlet combine to form the quintet of S_6.

The above example was completely resolved by group theoretical methods and no extra degeneracy occurred. Consider now the subgroup link given in part (b) of Table I. A spatial realization of the lower subgroup corresponds to vertices at the face center of a cube distorted along one axis joining opposite faces. Although the two intermediate subgroups are isomorphic to the cubic group they represent different embeddings in the group chain. Setting $f=d$ the matrix is invariant under S_4S_2 and the doublet and $([2]\otimes[1^2])$ $[2]$ singlet merge to form the $[3,1]$ $[2]$ triplet. Alternately setting $e=f$, $d=c$, and $b=a$ the matrix is invariant under S_2 wr S_3, the above radical factors giving roots $a+f+4c$ and $a+f-2c$ the latter of which merges with a singlet to form the $[2]\otimes[2,1]$ doublet. The doublet and the $([2]\otimes[2])$ $[1^2]$ singlet merge to form the $([2]\otimes[2],[1^2])$ triplet. Note by simply setting $a-e=b-f$ we have a five-parameter matrix with a triplet degeneracy which is invariant under neither of the higher subgroups. If the roots are arranged as above no crossing of levels to produce compatability with irrep species is required.

The extra degeneracy does not result from the multiplicity (i. e., from the value of the square root in the quadratic). No intermediate subgroup can exist in the decomposition S_2 wr S_3 $/(S_2$ wr $S_2)S_2$. This is another example of a systematic degeneracy which is not directly a result of group symmetry. This is similar to the extra degeneracy that may be obtained in the S_4S_2 symmetry by setting $a-e=b-d$ giving a quartet while no intermediate subgroup can exist in the chain S_6 $/S_4S_2$. Note also the radical factors by simply setting $a+e-b-f-2d=2c$.

Next consider the subgroup link of part (c) of Table I. The D_{2d} symmetry corresponds to the figure shown with the (2,3), (4,5), and (1,6) lines mutually perpendicular. If $a-f=b-c$ and $d-e=g-c$, the$[2]\otimes[1^2]$ radical factors giving roots $b+2g-3c$ and $b-a$; the latter of which merges with the doublet to form the $[3,1]$ $[2]$ triplet of S_4S_2. Note simply setting $d=e$ brings this matrix to the form invariant under $(S_2$ wr $S_2)S_2$ discussed above. A triplet degeneracy with no underlying group symmetry may be obtained in a manner similar to that discussed there.

A more complicated case involving the structure of triphenylmethyl with spatial substitution group S_2 wr S_3 yet showing fivefold degeneracies is discussed elsewhere.[6]

IV. CONCLUSIONS

Group theory is linked with degeneracy by the fact that if two operators S and T commute with an operator M then so does their product $[TS, M] = 0$, i.e., closure is required in both the operator realization and their matrix representation. However, matrix commutation $[T, M] = 0$ implies invariance under a similarity transformation $TMT^{-1} = M$ if and only if the inverse exists. The elements of a real symmetric matrix also form the basis of an $[(n+1)n/2]$-dimensional vector space and any substitutional symmetry among the matrix elements is also a substitutional symmetry operation in the vector space. The identification $(T^+MT)_{ij} = M_{kl}\ (T^*_{ki}T_{lj})$ shows that every unitary transformation of the matrix is a unitary transformation of the vector space. In particular, permutational (substitutional) transformations of the matrix give a permutational transformation of the vector space. But not every transformation of the vector space can be put in the factored form of a matrix transformation. In particular there exist substitutional transformations of the vector space which have no corresponding unitary transformation of the underlying matrix. Thus a matrix may show substitutional symmetries among its elements which do not correspond to a higher invariance group yet do result in greater degeneracy than that required by the actual invariance group.

The diagonalization of an Hermitian matrix may be considered in terms of perturbation theory. Suppose the set of n items divides naturally into two subsets n_1 and n_2 such that the invariant matrix has on its diagonal block matrices A_1 and A_2 describing the intraset interactions with substitutional symmetry groups G_1 and G_2, respectively. The interset interactions are represented by the off-diagonal $n_1 \times n_2$ block matrix A_{12} which in zeroth order is the null matrix. By setting all interset interactions to the same value the overall symmetry is not lowered. On reduction to irrep species the off-diagonal interset interaction matrix will have a nonzero entry only in the position connecting the totally symmetric irrep species of the two diagonal block matrices. Roots for the other irrep species are unaffected because the interset interaction matrix does not have the symmetry to effect them. In the special case $n = 2m$ the nonzero interset interaction lifts the zeroth-order degeneracy between the $[m] \otimes [2]$ and $[m] \otimes [1^2]$ modes. In Hückel models the interset interaction is usually limited to first neighbors in a specified spatial arrangement with all other interset matrix elements set equal to zero. This lowers the overall symmetry of the matrix to the point group allowed by the spatial arrangement. However, the intraset interactions in the diagonal blocks are usually considered as unmodified and therefore still represent the zeroth-order symmetry. When transformed to the irrep species of the zeroth-order symmetry group the interset interaction matrix often has vanishing entries for some irrep species. Thus some irrep species reflect the lower point group symmetry and degeneracies are lifted by the interset interactions while other irrep species are unaffected by the interset interactions and retain the degeneracies implicit in the zeroth-order symmetry group. In perturbation problems $H = H_0 + \lambda H_1$, where the zeroth-order symmetry G_0 is lowered by the perturbation to symmetry G, it is not uncommon for some zeroth-order degeneracies to persist through first-order perturbation calculations so that the model to this order has symmetry higher than G. Although analogous to this extra degeneracy in first-order perturbation theory, the Hückel model is slightly different in that the perturbation is completely ignored in the intraset interactions (the diagonal block matrices are not modified) and the matrix is diagonalized in a finite basis (several orders of perturbation in a truncated basis).

In summary we have outlined a group theoretical procedure for determining the general form of a real symmetric matrix invariant under a given substitutional symmetry group. The procedure gives the number of independent elements appearing in the invariant matrix and the degeneracies of the root form. Projection operators of the group are sufficient to block diagonalize the matrix within equivalent irreps. Further diagonalization involves the resolution of polynomials of order of the multiplicities which in general do not factor. By specific examples we have shown the degeneracies of the roots do not uniquely determine the substitutional symmetry group which leaves the matrix invariant. Substitutional symmetries that cannot be realized as similarity transformations forming a group can cause systematic degeneracies beyond those derivable from purely group theoretic analysis.

[1]W. Pauli, Z. Phys. **36**, 336 (1926).

[2]J. M. Jauch and E. L. Hill, Phys. Rev. **57**, 641 (1940).

[3]H. Ui and G. Takeda, Prog. Theor. Phys. **72**, 266 (1984); M. Moshinsky and C. Quesne, Ann. Phys. (NY) **148**, 462 (1983).

[4]U. Wild, J. Keller, and H. H. Gunthard, Theor. Chim. Acta **14**, 383 (1969); J. D. Roberts, A. Streitwibser, and C. M. Regan, J. Am. Chem. Soc. **74**, 4579 (1952).

[5]G. D. James and A. Kerber, *The Representation Theory of the Symmetric Group*, Vol. 16, Encyc. Math. and Appl. (Addison–Wesley, Reading, MA, 1981).

[6]T. H. Siddall and J. J. Sullivan, to be submitted to J. Chem. Phys.

Symmetries of linear systems of second-order ordinary differential equations

A. González López
School of Mathematics, 127 Vincent Hall, 216 Church Street S.E., University of Minnesota, Minneapolis, Minnesota 55455

(Received 20 March 1986; accepted for publication 22 December 1987)

In this paper, several problems concerning the Lie algebra structure of symmetries and variational symmetries of a general linear system of second-order ordinary differential equations are studied. In particular, a necessary and sufficient condition is obtained, in terms of the coefficients of the system, for the system's symmetry algebra to be of maximal dimension (i.e., n^2+4n+3) and isomorphic to $\mathrm{sl}(n+2,\mathbb{R})$, the well-known symmetry algebra of the free-particle equation $\mathbf{x}''=\mathbf{0}$. When this condition is satisfied, it is proved that the system is Lagrangian and that its variational symmetry algebra is isomorphic to a fixed, $(n^2+3n+6)/2$-dimensional Lie algebra, whose structure (Levi–Mal'cev decomposition and realization by means of a matrix algebra) is determined. For the particular case of *isotropic* systems (which includes, as far as is known, all the examples treated in the literature), explicit formulas for the generators of both the symmetry algebra and the variational symmetry algebra are obtained.

I. INTRODUCTION

Recently there has been a growing interest in the study of the continuous symmetries of systems of ordinary differential equations (ODE's), due to a wide range of reasons such as (i) the connection between continuous symmetries and first integrals of a system of ODE's,[1,2] even when the system is not derivable from a variational principle and Noether's classical theorem is not applicable; (ii) the valuable information about qualitative properties of the solutions of a system of ODE's that symmetries can sometimes provide[3,4]; and (iii) the recently conjectured importance of the structure of the symmetry group of a system of differential equations in connection with its quantization,[5] etc. As a consequence, the last decade has witnessed a significant extension of the number of differential equations whose Lie algebra of symmetry vectors is known. (A *symmetry vector* of a differential equation, called by Lie an *infinitesimal symmetry*, is just the generator of a one-parameter subgroup of symmetry transformations; it is a well-known fact[6] that the set of all the symmetry vectors of a differential equation forms a Lie algebra, whose associated Lie group is the group of all symmetry transformations continuously connected to the identity.) Such systems are, with a few exceptions,[7,8] second order and linear, and include the free-particle,[9] the harmonic oscillator with time-dependent frequency,[10] and the one-dimensional harmonic oscillator with constant damping.[11] It was found that the symmetry algebra of all these systems is not only of maximal dimension (i.e., n^2+4n+3), but it is also isomorphic to $\mathrm{sl}(n+2,\mathbb{R})$, n being the dimension of the system. (As is well-known, the dimension of the symmetry algebra of an analytic system of n second-order ODE's cannot be greater than n^2+4n+3; this result was established by Lie[9] for $n=1$, and was recently extended in Ref. 8 to systems of arbitrary dimension.)

It was conjectured on these grounds that the above is true for *all* linear (nonhomogeneous) systems of second-order differential equations. For single equations (i.e., $n=1$) this result was first established *locally* by Leach[12] using the Hamiltonian formalism. He showed that the restriction of any linear second-order differential equation to an appropriate open subset possesses an eight-dimensional symmetry algebra which is isomorphic to $\mathrm{sl}(3,\mathbb{R})$. This result was rederived by Martini and Kersten[13] by using Arnold's transformation (see the following section) to show that every linear second-order differential equation is locally equivalent to $x''=0$. These authors, however, did not obtain a closed formula for the generators of the symmetry algebra. Finally, Aguirre and Krause[14] showed by direct calculation that the symmetry algebra of every linear second-order differential equation is eight-dimensional. In contrast to Refs. 12 and 13, this was a *global* result. However, its authors, apparently unaware of the previous references, were not able to prove that the eight-dimensional symmetry algebra they found was isomorphic to $\mathrm{sl}(3,\mathbb{R})$. [This actually follows easily from a theorem of Lie according to which every second-order differential equation whose symmetry algebra is eight-dimensional is locally equivalent to $x''=0$, whose symmetry algebra is known to be isomorphic to $\mathrm{sl}(3,\mathbb{R})$.]

For systems of arbitrary dimension $n>1$, however, very little is known. Leach[12] conjectured in 1979 that the symmetry algebra of *every* linear system of n second-order differential equations is isomorphic to $\mathrm{sl}(n+2,\mathbb{R})$. He even advanced a heuristic argument justifying this claim for uncoupled, undamped, and homogeneous systems.[15]

We shall see in this paper that the above conjecture is actually false, even for the simple case just mentioned. This, of course, raises the problem of finding necessary and sufficient conditions in order that a linear second-order system,

$$\mathbf{x}''+2A_1(t)\cdot\mathbf{x}'+A_0(t)\cdot\mathbf{x}+\mathbf{b}(t)=\mathbf{0},\qquad \mathbf{x},\mathbf{b}\in\mathbb{R}^n,\quad A_0,A_1\ n\times n \text{ matrices},\tag{1.1}$$

be isomorphic to $\mathrm{sl}(n+2,\mathbb{R})$. This problem will be completely solved in Sec. II, where we shall establish the following result.

Theorem 1 (Main theorem): The symmetry algebra of the system (1.1) is isomorphic to $\mathrm{sl}(n+2,\mathbb{R})$ if and only if

 0022-2488/88/051097-09$02.50

there exists a scalar function a: $\mathbb{R} \to \mathbb{R}$ such that

$$A_0 = A_1' + A_1^2 + a\mathbf{1}. \tag{1.2}$$

In fact, we shall prove a stronger result, namely that (1.2) is also necessary and sufficient for the system (1.1) to admit a symmetry algebra of maximal dimension (i.e., $n^2 + 4n + 3$). In other words, among all $(n^2 + 4n + 3)$-dimensional Lie algebras *only* $\mathrm{sl}(n+2,\mathbb{R})$ can be isomorphic to the symmetry algebra of a linear second-order system. For $n = 1$, this result was proved by Lie (Ref. 9, p. 405) for arbitrary (not necessarily linear) equations using his famous classification of Lie algebras of vector fields in the plane.[16] To the best of our knowledge, however, the case $n > 1$ has not been dealt with at all in the literature. This is probably due to the fact that for $n > 2$, the classification of all Lie algebras of vector fields in $\mathbb{R}^n$ along the lines followed by Lie for $n = 2$ has proved to be a formidable task, so far accomplished only for $n \leqslant 6$.[17] Another consequence of the theorem is that (1.2) also characterizes all linear systems (1.1) which can be *locally* transformed to the canonical form

$$\frac{d^2\mathbf{y}}{du^2} = \mathbf{0} \tag{1.3}$$

by a suitable change of coordinates $(t,\mathbf{x}) \to (u,\mathbf{y})$, i.e., all linear second-order systems whose integral curves can be simultaneously rectified by a single change of local coordinates.

In Sec. III we shall show that all the linear systems (1.1) satisfying condition (1.2) (henceforth called *maximally symmetric*, for obvious reasons) are Lagrangian, i.e., they are equivalent to the system of Euler–Lagrange variational equations of some action

$$A = \int_{t_0}^{t_1} L(t,\mathbf{x},\mathbf{x}')\,dt. \tag{1.4}$$

We shall see that L is given explicitly by the formula

$$2L = R^{-1}(t)\,[\mathbf{x}' - \mathbf{x}_0'(t) - A_1(t)\cdot(\mathbf{x} - \mathbf{x}_0(t))]^2 - a(t)\,[R^{-1}(t)\cdot(\mathbf{x} - \mathbf{x}_0(t))]^2, \tag{1.5}$$

where $\mathbf{x}_0(t)$ is any particular solution of (1.1), $a(t)$ is defined by (1.2), and $R(t)$ is an $n \times n$ matrix satisfying

$$R' + A_1 R = 0, \quad R(0) \text{ nonsingular}. \tag{1.6}$$

We shall study the algebra of variational symmetry vectors of (1.4) and (1.5), proving that it is always isomorphic to a certain *fixed* Lie algebra g^V of dimension $(n^2 + 3n + 6)/2$. We shall also analyze the structure of this algebra, finding its Levi–Mal'cev decomposition[18] and an explicit realization of it as a matrix algebra. In particular, our results apply to such systems as the n-dimensional harmonic oscillator, the one-dimensional harmonic oscillator with time-dependent frequency, and the one-dimensional damped harmonic oscillator. The variational symmetry algebras of these systems have been computed by several authors[10,11] but, as far as we know, their structure was not analyzed; in particular, it was not even known if (for fixed n) all these algebras were isomorphic or not.

The method used to prove the above results relies on the simple fact that the property of being a symmetry vector of a system of differential equations is invariant under changes of coordinates. This allows us to simplify the problem at hand by a suitable choice of local coordinates, which in fact is a generalization of Arnold's transformation, already used in this context in Ref. 13 when $n = 1$. The results obtained in this way are, however, purely *local*. To obtain their *global* counterparts, we show that the local formulas we derive are actually globally well-defined. An outcome of this method of proof is an explicit expression for the generators of the symmetry algebra and of the variational symmetry algebra of an arbitrary *isotropic* linear system,

$$\mathbf{x}'' + a_1(t)\mathbf{x}' + a_0(t)\mathbf{x} + \mathbf{b}(t) = 0 \quad [a_0(t),a_1(t)\in\mathbb{R}], \tag{1.7}$$

in terms of its general solution. This formula provides a very simple way of computing the symmetry algebra and variational symmetry algebra of all the linear systems quoted above,[10,11] as we will show in Sec. IV with a practical example.

II. MAIN THEOREM

In this section we shall determine what are the necessary and sufficient conditions in order that the Lie algebra of symmetry vectors of the second-order linear system

$$\mathbf{x}'' + 2A_1(t)\cdot\mathbf{x}' + A_0(t)\cdot\mathbf{x} + \mathbf{b}(t) = \mathbf{0}, \qquad \mathbf{x},\mathbf{b}\in\mathbb{R}^n, \quad A_1,A_0 \ n\times n \text{ matrices}, \tag{2.1}$$

be isomorphic to $\mathrm{sl}(n+2,\mathbb{R})$. First of all, we can get rid of the second and fourth terms in the left-hand side of (2.1) by performing the (linear) change of variables

$$t = u, \quad \mathbf{x} = R(t)\mathbf{y} + \mathbf{x}_0(t), \tag{2.2}$$

where $\mathbf{x}_0(t)$ is again a particular solution of (2.1) and the $n\times n$ matrix $R(t)$ satisfies Eq. (1.6). Indeed, this change of variables transforms (2.1) into the system

$$\mathbf{y}'' + A(u)\cdot\mathbf{y} = \mathbf{0}, \tag{2.3}$$

where the matrix $A(u)$ is given by

$$A = R^{-1}(A_0 - A_1' - A_1^2)R. \tag{2.4}$$

[Notice that the invertibility of $R(t)$ for all t clearly follows from (1.6).] But the property of being a symmetry vector of a differential equation is invariant under changes of coordinates. In other words, if S is a symmetry vector of (2.3), then the push-forward of S under the *global* diffeomorphism (2.2) will be a symmetry vector of (2.1).[19] This implies that *the symmetry algebras of (2.1) and (2.3) are isomorphic*, since push-forwards preserve the Lie–Jacobi bracket, i.e.,

$$\Psi_*[X,Y] = [\Psi_*X,\Psi_*Y], \tag{2.5}$$

if X,Y are vector fields, Ψ is a diffeomorphism, and Ψ_* denotes its push-forward.[20]

Next, we shall find a necessary condition for the symmetry algebra of (2.3) to be of dimension $n^2 + 4n + 3$.

Proposition 1: If the symmetry algebra of the system (2.3) has dimension $n^2 + 4n + 3$, then $A(u)$ is a multiple of the unit matrix, i.e.,

$$A(u) = a(u)\mathbf{1}, \quad \text{all } u\in\mathbb{R}, \tag{2.6}$$

for some scalar function $a(u)$.

Proof: If $n = 1$ there is nothing to prove, so we shall assume that $n \geqslant 2$ in what follows. The necessary and sufficient condition in order that the vector field

$$S = \theta(u,\mathbf{y})\partial_u + \boldsymbol{\xi}(u,\mathbf{y})\partial_{\mathbf{y}} \tag{2.7}$$

be a symmetry vector of (2.3) is[6,8] that

$$\boldsymbol{\xi}^{(2)}(u,\mathbf{y},\mathbf{y}', -A(u)\cdot\mathbf{y}) + A(u)\cdot\boldsymbol{\xi}(u,\mathbf{y}) + \theta(u,\mathbf{y})A'(u)\cdot\mathbf{y} = 0 \tag{2.8}$$

holds identically in $(u,\mathbf{y},\mathbf{y}')$, where $\boldsymbol{\xi}^{(2)} \equiv (\xi_1^{(2)},...,\xi_n^{(2)})$ is given by

$$\xi_i^{(2)} = \xi_{i,uu} + \sum_{1\leqslant j\leqslant n}(2\xi_{i,ju} - \delta_{ij}\theta_{uu})y_j' + \sum_{1\leqslant j,k\leqslant n}(\xi_{i,jk} - 2\theta_{ju}\delta_{ik})y_j'y_k' - \sum_{1\leqslant j,k\leqslant n}\theta_{jk}y_i'y_j'y_k' + \sum_{1\leqslant j\leqslant n}\left[\xi_{i,j} - y_i'\theta_j - 2\delta_{ij}\left(\theta_u + \sum_{1\leqslant k\leqslant n}\theta_k y_k'\right)\right]y_j''. \tag{2.9}$$

(The subindices $u,j,k,...$ in θ and ξ_i denote partial differentiation with respect to the variables $u, y_j, y_k,...$.) Substituting (2.9) into (2.8) and equating to zero the coefficients of $1, y_j'$, $y_j'y_k'$, and $y_i'y_j'y_k'$ in the resulting identity, we arrive at the following system of partial differential equations in ξ and θ:

$$\xi_{i,uu} - \sum_{1\leqslant j,k\leqslant n}A_{jk}\xi_{i,j}\,y_k + \sum_{1\leqslant k\leqslant n}(2\theta_u A_{ik}\,y_k + A_{ik}\xi_k + \theta A'_{ik}\,y_k) = 0, \tag{2.10}$$

$$2\xi_{i,ju} + \delta_{ij}\left(\sum_{1\leqslant l,k\leqslant n}\theta_l A_{lk}\,y_k - \theta_{uu}\right) + 2\sum_{1\leqslant k\leqslant n}\theta_j A_{ik}\,y_k = 0, \tag{2.11}$$

$$\xi_{i,jk} - \theta_{ju}\delta_{ik} - \theta_{ku}\delta_{ij} = 0, \tag{2.12}$$

$$\theta_{jk} = 0, \quad 1\leqslant i,j,k\leqslant n, \tag{2.13}$$

where $(A_{ij})_{1\leqslant i,j\leqslant n} = A$. We claim that if A is *not* of the form (2.6), then these equations necessarily imply that

$$\theta_j = 0, \quad 1\leqslant j\leqslant n. \tag{2.14}$$

Indeed, differentiating (2.11) with respect to u and using (2.13), we get the identity

$$\delta_{ij}\left(\theta_{kuu} + \sum_{1\leqslant l\leqslant n}\theta_l A_{lk}\right) + 2\delta_{ik}\theta_{juu} + 2\theta_j A_{ik} = 0, \quad 1\leqslant i,j,k\leqslant n. \tag{2.15}$$

If A is not proportional to the unit matrix, then either there exist $u_0\in\mathbb{R}$ and $p\neq q$ such that $A_{pq}(u_0)\neq 0$, or else

$$A_{ij}(u) = \delta_{ij}a_i(u), \quad \text{all } u\in\mathbb{R}, \quad 1\leqslant i,j\leqslant n, \tag{2.16}$$

with $a_p(u_0)\neq a_q(u_0)$. By continuity, in the first case there would be an open interval $I\ni t_0$ such that

$$A_{pq}(u)\neq 0, \quad \text{all } u\in I, \tag{2.17}$$

whereas in the second case there would be an open interval $I\ni u_0$ such that

$$a_p(u)\neq a_q(u), \quad \text{all } u\in I. \tag{2.18}$$

Furthermore, it follows from (2.13) that θ_j is a function of u only. Suppose now that we are in the first case, i.e., that (2.17) holds for some $p\neq q$. We then let $i=p$, $k=q$ in (2.15), obtaining

$$\theta_j A_{pq} = 0, \quad \text{all } j\neq p, \tag{2.19}$$

whence, from (2.17),

$$\theta_j(u) = 0, \quad \text{all } u\in I, \quad \text{all } j\neq p. \tag{2.20}$$

Using now (2.15) with $i=j=p$ and $k=q$, and taking (2.20) into account, we obtain

$$\theta_j(u) = 0, \quad \text{all } u\in I, \quad 1\leqslant j\leqslant n. \tag{2.21}$$

Similarly, if we are in the second case, i.e., (2.16)–(2.18) hold, from (2.15) we obtain

$$\delta_{ij}(\theta_{kuu} + a_k\theta_k) + 2\delta_{ik}(\theta_{juu} + a_i\theta_j) = 0, \quad 1\leqslant i,j,k\leqslant n. \tag{2.22}$$

Letting first $k=i\neq j$ and then $i=j\neq k$ in this identity, we easily get

$$\theta_{juu} + a_i\theta_j = 0, \quad 1\leqslant i,j\leqslant n, \tag{2.23}$$

from which it follows that

$$(a_p - a_q)\theta_j = 0, \quad 1\leqslant i,j\leqslant n, \tag{2.24}$$

whence, taking into account (2.18), (2.21) follows again. Finally, we observe that the functions $\theta_j(u)$ $(j=1,...,n)$ satisfy the linear system of ODE's

$$(\theta_j)'' + A_{i(j)i(j)}(u)\theta_j = 0, \quad 1\leqslant j\leqslant n, \tag{2.25}$$

obtained from (2.15) by choosing, for every $j\in\{1,...,n\}$, a fixed $i(j)\in\{1,...,n\}$ different from j, and setting $i=k=i(j)$. From (2.21) and (2.25) it then follows that all the θ_j's vanish identically, as claimed.

We will now show that (2.14) implies that the general solution of the system (2.10)–(2.13) depends on less than n^2+4n+3 arbitrary constants. Indeed, substituting (2.14) into (2.11)–(2.13) we obtain

$$2\xi_{i,ju} - \delta_{ij}\theta_{uu} = 0, \tag{2.26}$$

$$\xi_{i,jk} = 0, \quad 1\leqslant i,j,k\leqslant n. \tag{2.27}$$

From these equations it follows that ξ_i has the structure

$$\xi_i = \theta'(u)y_i/2 + \sum_{1\leqslant j\leqslant n}b_{ij}y_j + C_i(u), \quad 1\leqslant i\leqslant n, \tag{2.28}$$

where the b_{ij}'s are real constants. Inserting this into (2.10) and equating to zero the coefficients of 1 and y_k in the resulting identity, we finally get

$$\theta'''\delta_{ik} + 4\theta'A_{ik} + 2\theta A'_{ik} + 2\sum_{1\leqslant i\leqslant n}(A_{ij}b_{jk} - b_{ij}A_{jk}) = 0, \tag{2.29}$$

$$C_i'' + \sum_{1\leqslant k\leqslant n}A_{ik}C_k = 0, \quad 1\leqslant i,k\leqslant n. \tag{2.30}$$

If $\gamma_i(u) \equiv (\gamma_{i1}(u),...,\gamma_{in}(u))$ $(i=1,2)$ is a fixed basis of solutions of the linear system (2.30), then we have

$$C_i(u) = c_{1i}\gamma_{1i}(u) + c_{2i}\gamma_{2i}(u), \quad 1\leqslant i\leqslant n, \tag{2.31}$$

where c_{1i}, c_{2i} $(1\leqslant i\leqslant n)$ are $2n$ arbitrary real constants. Similarly, taking $i=k=1$ in (2.29), it follows that

$$\theta(u) = d_1\Theta_1(u) + d_2\Theta_2(u) + d_3\Theta(u) + \sum_{2\leqslant i\leqslant n}[b_{j1}\Phi_{1j}(u) - b_{1j}\Phi_{j1}(u)], \tag{2.32}$$

where Θ_1, Θ_2, Θ_3 are a basis of solutions of the homogeneous linear equation

$$\theta''' + 4\theta' A_{11} + 2\theta A'_{11} = 0, \tag{2.33}$$

where Φ_{ij} denotes a particular solution of the linear equation

$$\theta''' + 4\theta' A_{11} + 2\theta A'_{11} + 2A_{ij} = 0, \tag{2.34}$$

and where d_1, d_3 are arbitrary real constants. From (2.28), (2.31), and (2.32) it follows that when A is not of the form (2.6), the general solution of the system of PDE's (2.10)–(2.13) depends on *at most* $n^2 + 2n + 3$ arbitrary constants d_i, c_{1i}, c_{2i}, and b_{ij} $(1 \leqslant i,j \leqslant n)$. [It will generally depend on *less* than $n^2 + 2n + 3$ arbitrary constants since, substituting (2.32) into (2.29) for $(i,k) \neq (1,1)$, we will in general obtain several linear relations between the d_i's and the b_{ij}'s]. This implies that when A is not of the form (2.6), the symmetry algebra of (2.3) has dimension less than or equal to $n^2 + 2n + 3$, which is less than $n^2 + 4n + 3$. This concludes the proof of the proposition.

We shall now prove that when condition (2.6) is satisfied, then the symmetry algebra of the system (2.3) is isomorphic to $\mathrm{sl}(n+2,\mathbb{R})$. To do this, we shall consider a system slightly more general than (2.3)–(2.6), i.e., the general *isotropic* system

$$\mathbf{x}'' + a_1(t)\mathbf{x}' + a_0(t)\mathbf{x} + \mathbf{b}(t) = \mathbf{0}, \tag{2.35}$$

where a_0 and a_1 are both scalar functions (notice that we have called the dependent and independent variables t and $\mathbf{x}$ again; this should not confuse the reader.) Let $\mathbf{x}_0(t)$ be any particular solution of this system, and denote by $x_1(t)$ and $x_2(t)$ a basis of solutions of the associated homogeneous scalar ODE

$$x'' + a_1(t)x' + a_0(t)x = 0 \quad (x \in \mathbb{R}). \tag{2.36}$$

The general solution of (2.36) can then be expressed as follows:

$$\mathbf{x} = \mathbf{x}_0(t) + \mathbf{c}_1 x_1(t) + \mathbf{c}_2 x_2(t), \tag{2.37}$$

where $\mathbf{c}_1, \mathbf{c}_2 \in \mathbb{R}^n$ are constant. We can rewrite (2.37) as

$$[\mathbf{x} - \mathbf{x}_0(t)]/x_1(t) = \mathbf{c}_1 + \mathbf{c}_2[x_2(t)/x_1(t)], \quad \text{all } t \in J, \tag{2.38}$$

where

$$J = \{t \in \mathbb{R} \mid x_1(t) \neq 0\}. \tag{2.39}$$

Let us now perform the local change of coordinates

$$\Phi\colon V \equiv J \times \mathbb{R}^n \to W \equiv \Phi(J \times \mathbb{R}^n),$$

defined as follows:

$$u = x_2(t)/x_1(t), \quad \mathbf{y} = [\mathbf{x} - \mathbf{x}_0(t)]/x_1(t), \quad (t,\mathbf{x}) \in V, \tag{2.40}$$

which is a generalization of Arnold's transform.[19] Then (2.37) is transformed into

$$\mathbf{y} = \mathbf{c}_1 + \mathbf{c}_2 u \quad [(u,\mathbf{y}) \in W], \tag{2.41}$$

which is the general solution of the system (1.3). Hence (2.40) transforms the restriction of (2.35) to the open subset V into (1.3) restricted to W. It follows [see the remark after (2.4)] that the symmetry algebra of these two equations are isomorphic under Φ_*. The symmetry algebra of (1.3) is known to be isomorphic to $\mathrm{sl}(n+2,\mathbb{R})$; a basis of it is given by the following $n^2 + 4n + 3$ vector fields[8]:

$$\begin{aligned} &Y_1 = \partial_u, \quad Y_2 = u\,\partial_u, \quad Y_3^i = y_i\,\partial_u, \quad Y_4^i = \partial_i, \\ &Y_5^i = u\,\partial_i, \quad Y_6^{ij} = y_i\,\partial_j, \quad Y_7 = u(u\,\partial_u + \mathbf{y}\,\partial_\mathbf{y}), \\ &Y_8^i = y_i(u\,\partial_u + \mathbf{y}\,\partial_\mathbf{y}), \quad 1 \leqslant i,j \leqslant n, \quad \partial_\mathbf{y} \equiv (\partial_1,\ldots,\partial_n). \end{aligned} \tag{2.42}$$

Applying $(\Phi^{-1})_*$ to the vector fields (2.42) we obtain the following basis of the symmetry algebra of the restriction of (2.35) to V:

$$\begin{aligned} X_\alpha &= W(t)^{-1} x_\alpha(t)\,[x_1(t)\partial_t + (x_1'(t)\mathbf{x} - \mathbf{W}_1(t))\partial_\mathbf{x}]; \\ X_3^i &= W(t)^{-1}(x_i - x_{0i}(t)) \\ &\quad \times [x_1(t)\partial_t + (x_1'(t)\mathbf{x} - \mathbf{W}_1(t))\partial_\mathbf{x}]; \\ X_4^i &= x_1(t)\partial_i; \quad X_5^i = x_2(t)\partial_i; \quad X_6^{ij} = (x_i - x_{0i}(t))\partial_j; \\ X_7 &= W(t)^{-1} x_2(t)\,[x_2(t)\partial_t + (x_2'(t)\mathbf{x} - \mathbf{W}_2(t))\partial_\mathbf{x}]; \\ X_8^i &= W(t)^{-1}(x_i - x_{0i}(t)) \\ &\quad \times [x_2(t)\partial_t + (x_2'(t)\mathbf{x} - \mathbf{W}_2(t))\partial_\mathbf{x}]; \\ &\quad 1 \leqslant i,j \leqslant n, \quad \alpha = 1,2, \quad \partial_\mathbf{x} \equiv (\partial_1,\ldots,\partial_n), \end{aligned} \tag{2.43}$$

where $(t,\mathbf{x}) \in V$, and we have used the notation

$$\begin{aligned} &\mathbf{x}_0(t) \equiv (x_{01}(t),\ldots,x_{0n}(t)), \\ &W = x_1 x_2' - x_1' x_2, \quad \mathbf{W}_i = \mathbf{x}_0 x_i' - \mathbf{x}_0' x_i. \end{aligned} \tag{2.44}$$

We should also remark that the bases (2.42) and (2.43) have been labeled in such a way that

$$X_\alpha = (\Phi^{-1})_* Y_\alpha, \tag{2.45}$$

where α stands for the various subindices and superindices labeling (2.42) and (2.43). From the way we have constructed them, the vector fields X_α are *a priori* defined only locally (i.e., on V). However, a glance at the explicit formula (2.43) shows that the X_α's are *well-defined on all of* $\mathbb{R} \times \mathbb{R}^n$, since the Wronskian of two linearly independent solutions of a linear ODE never vanishes. A simple continuity argument then proves that these vector fields are symmetry vectors of (2.35). By (2.5) and (2.45), we also know that the restrictions of (2.43) to V form a Lie algebra *with the same structure constants as (2.42).* Again by continuity, it follows that this is also true globally. Finally, suppose that S is a symmetry vector of (2.35). The restriction of S to V is then a symmetry vector of (2.35) restricted to V, and therefore it is spanned by the restrictions of (2.43) to V, i.e.,

$$S|_V = \sum_{1 \leqslant \alpha \leqslant N} c_\alpha X_\alpha|_V \quad (N \equiv n^2 + 4n + 3). \tag{2.46}$$

Since, by definition of J, $V \equiv J \times \mathbb{R}^n$ is dense in $\mathbb{R} \times \mathbb{R}^n$, from (2.46) it follows that S is in the span of the X_α's. Hence (2.43) generates the symmetry algebra of (2.35). This proves the following proposition.

Proposition 2: The vector fields (2.43) are a basis of the symmetry algebra of the system (2.35) having the same structure constants as the standard basis (2.42) of $d^2\mathbf{y}/du^2 = 0$. In particular, the symmetry algebra of (2.35) is $(n^2 + 4n + 3)$-dimensional and isomorphic to $\mathrm{sl}(n+2,\mathbb{R})$.

For future reference, we shall list the commutation relations satisfied by the generators (2.43) [or, equivalently, (2.42)] in Table I.

The above proposition provides a simple way of com-

TABLE I. The symmetry algebra of a linear second-order system.

	X_1	X_2	X_3^k	X_4^k	X_5^k	X_6^{kl}	X_7	X_8^k
X_1	0	X_1	0	0	X_4^k	0	$2X_2 + \operatorname{tr} X_6$	X_3^k
X_2	$-X_1$	0	$-X_3^k$	0	X_5^k	0	X_7	0
X_3^i	0	X_3^i	0	$-\delta_{ik}X_1$	$X_6^{ik} - \delta_{ik}X_2$	$-\delta_{il}X_3^k$	X_8^i	0
X_4^i	0	0	$\delta_{ik}X_1$	0	0	$\delta_{il}X_4^l$	X_5^i	$\delta_{ik}(X_2 + \operatorname{tr} X_6) + X_6^{ki}$
X_5^i	$-X_4^i$	$-X_5^i$	$\delta_{ik}X_2 - X_6^{ik}$	0	0	$\delta_{ik}X_5^l$	0	$\delta_{ik}X_7$
X_6^{ij}	0	0	$\delta_{jk}X_3^i$	$-\delta_{ik}X_5^j$	$\delta_{jk}X_3^i$	$\delta_{jk}X_6^{il} - \delta_{il}X_6^{kj}$	0	$\delta_{jk}X_8^i$
X_7	$-(2X_2 + \operatorname{tr} X_6)$	$-X_7$	$-X_8^k$	$-X_5^k$	0	0	0	0
X_8^i	$-X_3^i$	0	0	$-\delta_{ik}(X_2 + \operatorname{tr} X_6) - X_6^{ki}$	$-\delta_{ik}X_7$	$-\delta_{il}X_8^k$	0	0

puting the symmetry algebra of the isotropic linear system (2.35) when its general solution is known (as is the case with all the systems quoted in the Introduction!). We shall illustrate this point in Sec. IV with a practical example.

Since the system (2.3)–(2.6) is clearly isotropic, from Propositions 1 and 2 we obtain this additional proposition.

Proposition 3: The necessary and sufficient condition for the symmetry algebra of the system (2.3) to be $(n^2 + 4n + 3)$-dimensional is that (2.6) holds. Moreover, when this condition is satisfied the symmetry algebra of (2.3) is isomorphic to $\mathrm{sl}(n + 2,\mathbf{R})$.

Finally, this proposition yields easily the main theorem of this section.

Theorem 2: The symmetry algebra of the linear second-order system (2.1) is $(n^2 + 4n + 3)$-dimensional if and only if there is a scalar function a: $\mathbf{R} \to \mathbf{R}$ such that

$$A_0 = A_1' + A_1^2 + a\mathbf{1}. \tag{2.47}$$

When this is the case, the symmetry algebra of (2.1) is isomorphic to $\mathrm{sl}(n + 2,\mathbf{R})$.

Proof: We have shown above that (2.1), (2.3), and (2.4) are equivalent under the diffeomorphism (2.2); therefore their symmetry algebras are isomorphic. The theorem then follows from Proposition 3 and (2.4).

Notice, in particular, that from (2.47) it follows that it is neither necessary nor sufficient that (2.1) be uncoupled (i.e., A_0 and A_1 diagonal), undamped (i.e., $A_1 = 0$), or homogeneous (i.e., $\mathbf{b} = \mathbf{0}$) for the symmetry algebra of (2.1) to be isomorphic to $\mathrm{sl}(n + 2,\mathbf{R})$. A counterexample is provided, for instance, by the system

$$x_i'' + \omega_i^2 x_i = 0, \quad 1 \leqslant i \leqslant n, \quad n > 1, \quad \omega_i \neq \omega_j, \quad \text{for } i \neq j. \tag{2.48}$$

[The symmetry algebra of this system can be explicitly computed and turns out to be $(3n + 1)$-dimensional.] This refutes the conjecture advanced in Ref. 15.

Before closing this section, we would like to mention a geometric consequence of Theorem 2. By a well-known result of Lie (Ref. 9, p. 405), a (not necessarily linear) second-order ODE can be locally transformed into the equation $x'' = 0$ by a change of dependent and independent variables if and only if its symmetry algebra is isomorphic to $\mathrm{sl}(3,\mathbf{R})$. For general systems of second-order ODE's, the author of the present paper is not aware of any similar result. However, from Theorem 2 (and its proof) it immediately follows that Lie's result is true for *linear* systems of second-order ODE's.

Corollary 1: The system (2.1) can be locally transformed to the form $d^2\mathbf{y}/du^2 = \mathbf{0}$ by a suitable change of variables $(t,\mathbf{x}) \to (u,\mathbf{y})$ if and only if condition (2.47) holds locally.

Proof: The symmetry algebra of $d^2\mathbf{y}/du^2 = \mathbf{0}$, and of any restriction of this system to an open subset, is known to be isomorphic to $\mathrm{sl}(n + 2,\mathbf{R})$. Since equivalent systems of differential equations possess isomorphic symmetry algebras, it follows that a necessary condition for (2.1) (possibly restricted to an open subset V) to be equivalent to $d^2\mathbf{y}/du^2 = \mathbf{0}$ is that its symmetry algebra be isomorphic to $\mathrm{sl}(n + 2,\mathbf{R})$. By Theorem 2, this implies that condition (2.47) must be satisfied (at least locally). Conversely, if (2.47) is satisfied in some open subset $J \subset \mathbf{R}$, then (2.2) (restricted to $J \times \mathbf{R}^n$) transforms (2.1) into $\mathbf{y}'' + a(u)\mathbf{y} = \mathbf{0}$, which is isotropic and therefore equivalent to $d^2\mathbf{y}/du^2 = \mathbf{0}$ under Arnold's transformation.

III. VARIATIONAL SYMMETRIES

In this section we shall study the variational symmetries of *maximally symmetric* second-order linear systems. By definition, these are the second-order linear systems whose symmetry algebra is of maximal dimension, i.e., $n^2 + 4n + 3$; according to Theorem 2 of Sec. II, they are characterized by condition (2.47). An important example is provided by isotropic systems:

$$\mathbf{x}'' + a_1(t)\mathbf{x}' + a_0(t)\mathbf{x} + \mathbf{b}(t) = \mathbf{0},$$
$$\mathbf{x},\mathbf{b} \in \mathbf{R}^n, \quad a_0(t), a_1(t) \in \mathbf{R}. \tag{3.1}$$

As remarked above, all the linear systems whose variational symmetry algebras have been computed in the references quoted in the Introduction are of this form. We shall prove below that all maximally symmetric linear systems are Lagrangian, and shall find the structure of their variational symmetry algebras.

Let us start by establishing our notation and quoting a few well-known results that will be useful in the sequel. Two functions f,g: $\mathbf{R}^m \to \mathbf{R}^n$ will be called *equivalent* if they have the same set of zeros, i.e., if $f(\mathbf{z}) = 0 \Leftrightarrow g(\mathbf{z}) = 0$, for all $\mathbf{z} \in \mathbf{R}^m$. A second-order system

$$\mathbf{x}'' = F(t,\mathbf{x},\mathbf{x}') \tag{3.2}$$

is Lagrangian if there exists a function $L(t,\mathbf{x},\mathbf{x}')$, called a Lagrangian for the system, such that $\mathbf{E}(L)$ is equivalent to $\mathbf{x}'' - \mathbf{F}(t,\mathbf{x},\mathbf{x}')$; here $\mathbf{E}$ is the Euler–Lagrange operator, defined by

$$\mathbf{E} = \partial_{\mathbf{x}} - D_t\, \partial_{\mathbf{x}'}, \tag{3.3}$$

where

$$D_t = \partial_t + \mathbf{x}'\, \partial_{\mathbf{x}} + \mathbf{x}''\, \partial'_{\mathbf{x}} \tag{3.4}$$

is the total derivative with respect to the independent variable t. In other words, (3.2) and $\mathbf{E}(L) = 0$ define the same system of differential equations. It should be noticed that (3.2) can be Lagrangian without $\mathbf{x}'' - \mathbf{F}(t,\mathbf{x},\mathbf{x}')$ being of the form $\mathbf{E}(f)$ for any $f(t,\mathbf{x},\mathbf{x}')$.[21] Given a function $L(t,\mathbf{x},\mathbf{x}')$, we define its Cartan one-form θ_L as follows:

$$\theta_L = L\, dt + \frac{\partial L}{\partial \mathbf{x}'}\, \boldsymbol{\omega}, \tag{3.5}$$

where $\boldsymbol{\omega} \equiv (\omega_1,\dots,\omega_n)$ is the vector-valued contact one-form

$$\boldsymbol{\omega} = d\mathbf{x} - \mathbf{x}'\, dt. \tag{3.6}$$

If (3.2) is Lagrangian and if L is a Lagrangian for (3.2), we say that a vector field S is a *variational symmetry (vector) of (3.2) relative to the Lagrangian L* if S is a variational symmetry vector of the action

$$A[\mathbf{x}] = \int_{t_1}^{t_0} L(t,\mathbf{x},\mathbf{x}')dt, \tag{3.7}$$

as defined in Ref. 6. It is well-known[22] that a necessary and sufficient condition for this is that S satisfies

$$\mathscr{L}_{S^{(1)}}\theta_L = df \tag{3.8}$$

for some function $f(t,\mathbf{x})$. Here $\mathscr{L}_{S^{(1)}}$ denotes the Lie derivative along the vector field $S^{(1)}$, where $S^{(1)}$ is the *first prolongation* of S, defined as follows[6]:

$$S^{(1)} = S + (D_t\boldsymbol{\xi} - \mathbf{x}'D_t\tau)\partial_{\mathbf{x}'}$$
$$[S \equiv \tau(t,\mathbf{x})\partial_t + \boldsymbol{\xi}(t,\mathbf{x})\partial_{\mathbf{x}}]. \tag{3.9}$$

The set of variational symmetries of a Lagrangian system relative to a Lagrangian L is a Lie subalgebra of the Lie algebra of *all* its symmetry vectors.[6] We shall call this subalgebra the *variational symmetry algebra* of the system relative to L. Noether's theorem[6,22] states that to every variational symmetry S of the action (3.7) there corresponds a first integral I_s of $\mathbf{E}(L) = 0$ given by

$$I_S = f - \langle S^{(1)},\theta_L\rangle, \tag{3.10}$$

where $\langle\ ,\ \rangle$ is the natural pairing between vector fields and one-forms.

Remark: It should be noted that the system (3.2) can be Lagrangian with respect to two *nonequivalent* Lagrangians L_1 and L_2 (i.e., $L_2 \neq cL_1 + D_t f$, $c = \text{const}$). When this is the case, the variational symmetry algebras of (3.2) relative to L_1 and L_2 need not be the same, nor even isomorphic. An example of this statement is provided by the system $\mathbf{x}'' = \mathbf{0}$ and the Lagrangians $L_1 = \mathbf{x}'^2/2$ and $L_2 = \exp L_1$. We leave the details to the reader.

We shall now quote a standard result about the behavior of Lagrangian systems under a diffeomorphism

$$\Phi = (\varphi_0,\boldsymbol{\varphi}): A \subset \mathbb{R}\times\mathbb{R}^n \to B \subset \mathbb{R}\times\mathbb{R}^n.$$

Let

$$A' = \left\{(t,\mathbf{x},\mathbf{x}')\in A\times\mathbb{R}^n \,\middle|\, \frac{\partial\varphi_0}{\partial t}(t,\mathbf{x}) + \mathbf{x}'\frac{\partial\varphi_0}{\partial\mathbf{x}}(t,\mathbf{x}) \neq 0\right\}, \tag{3.11}$$

and define the *first prolongation* $\Phi^{(1)}: A' \to B\times\mathbb{R}^n$ of Φ as follows:

$$\Phi^{(1)}(t,\mathbf{x},\mathbf{x}') = (\Phi(t,x),(\boldsymbol{\varphi}'/\varphi'_0)(t,\mathbf{x},\mathbf{x}')), \tag{3.12}$$

where $f' \equiv D_t f$. Then $\Phi^{(1)}$ is also a diffeomorphism onto its image $B' = \Phi^{(1)}(A') \subset B\times\mathbb{R}^n$. Suppose now that the second-order systems

$$\mathbf{x}'' = \mathbf{F}_A(t,\mathbf{x},\mathbf{x}'), \quad (t,\mathbf{x},\mathbf{x}')\in A', \tag{3.13}$$

and

$$y'' = \mathbf{F}_B(u,\mathbf{y},\mathbf{y}'), \quad (u,\mathbf{y},\mathbf{y}')\in B', \tag{3.14}$$

are transformed into one another by the diffeomorphism Φ, and that (3.14) is Lagrangian. Let $L_B(u,\mathbf{y},\mathbf{y}')$ be a Lagrangian for (3.14), and define

$$L_A = (L_B\circ\Phi^{(1)})\varphi'_0. \tag{3.15}$$

Then we have the following lemma.

Lemma 1: The system (3.13) is also Lagrangian, and (3.15) is a Lagrangian for (3.13). Furthermore, if S_A is a variational symmetry of (3.13) relative to the Lagrangian (3.15), then

$$S_B = \Phi_* S_A \tag{3.16}$$

is a variational symmetry of (3.14) relative to L_B. Therefore the variational symmetry algebras of (3.14) and (3.13) relative to L_B and L_A are *isomorphic* under Φ_*. Finally, if I_A and I_B are the first integrals associated to S_A and S_B, respectively, by Noether's theorem, then we have

$$I_A = I_B\circ\Phi^{(1)}. \tag{3.17}$$

We shall begin by considering the variational symmetry algebra of the system $d^2\mathbf{y}/du^2 = \mathbf{0}$ relative to the standard Lagrangian

$$L_B = \mathbf{y}'^2/2. \tag{3.18}$$

A straightforward calculation shows that the following $(n^2+3n+6)/2$ vector fields are a basis for the variational symmetry algebra of this system:

$$V_1 = Y_1,\quad V_2 = 2Y_2 + \sum_{1\leqslant i\leqslant n} Y_6^{ii},\quad V_3 = Y_7,\quad V_4^i = Y_4^i,$$
$$V_5^i = Y_5^i,\quad V_6^{ij} = Y_6^{ij} - Y_6^{ji},\quad 1\leqslant i<j\leqslant n, \tag{3.19}$$

where the Y's are the generators (2.42) of the complete symmetry algebra of $\mathbf{y}'' = \mathbf{0}$. The first integrals associated to this basis by Noether's theorem are

$$J_1 = \mathbf{y}'^2/2,\quad J_2 = u\mathbf{y}'^2 - \mathbf{y}\mathbf{y}',\quad J_3 = (\mathbf{y} - u\mathbf{y}')^2/2,$$
$$J_4^i = -y_i',\quad J_5^i = y_i - uy_i',\quad J_6^{ij} = y_i'y_j - y_iy_j',$$
$$1\leqslant i<j\leqslant n. \tag{3.20}$$

They are functionally dependent on the $2n$ independent first integrals $\mathbf{J}_4 \equiv (J_4^1,\dots,J_4^n)$ and $\mathbf{J}_5 \equiv (J_5^1,\dots,J_5^n)$, since

$$J_1 = J_4^2/2,\quad J_2 = \mathbf{J}_4\mathbf{J}_5,\quad J_3 = J_5^2/2,$$
$$J_6^{ij} = J_4^iJ_5^j - J_4^jJ_5^i. \tag{3.21}$$

But we know from Sec. II that the general isotropic system (3.1) is locally equivalent to $\mathbf{y}''=\mathbf{0}$ under Arnold's mapping (2.40). Therefore we can apply Lemma 1 to these systems, with Φ given by (2.40) and A replaced by the open subset $V=J\times\mathbf{R}^n$ defined in Sec. II [see (2.39)]. An easy calculation then shows that

$$L_A=\frac{[x_1(t)\mathbf{x}'-x_1'(t)\mathbf{x}-\mathbf{W}_1(t)]^2}{2x_1^2(t)W(t)}, \tag{3.22}$$

where the notation was explained in Sec. II. By (2.45) and Lemma 1, the vector fields

$$N_1=X_1,\quad N_2=2X_2+\sum_{1\leqslant i\leqslant n}X_6^{ii},\quad N_3=X_7,\quad N_4^i=X_4^i,$$

$$N_5^i=X_5^i,\quad N_6^{ij}=X_6^{ij}-X_6^{ji},\quad 1\leqslant i<j\leqslant n, \tag{3.23}$$

are a basis of the variational symmetry algebra of the restriction of (3.1) to $V'=V\times\mathbf{R}^n$ relative to (3.22), the X's being defined in (2.43). Using (2.43) in (3.23), we get the explicit formulas

$$\begin{aligned}
N_1&=W(t)^{-1}x_1(t)[x_1(t)\partial_t+(x_1'(t)\mathbf{x}-\mathbf{W}(t))\partial_\mathbf{x}],\\
N_2&=W(t)^{-1}[2x_1(t)x_2(t)\partial_t+((x_1'(t)x_2(t)\\
&\qquad+x_1(t)x_2'(t))\mathbf{x}-\mathbf{W}_{12}(t))\partial_\mathbf{x}],\\
N_3&=W(t)^{-1}x_2(t)[x_2(t)\partial_t+(x_2'(t)\mathbf{x}-\mathbf{W}(t))\partial_\mathbf{x}],\\
N_4^i&=x_1(t)\partial_i,\quad N_5^i=x_2(t)\partial_i,\\
N_6^{ij}&=(x_i-x_{0i}(t))\partial_i-(x_j-x_{0j}(t))\partial_j,\quad 1\leqslant i<j\leqslant n,
\end{aligned} \tag{3.24}$$

where $t\in J$ and

$$\mathbf{W}_{12}\equiv\mathbf{x}_0(x_1x_2)'-\mathbf{x}_0'x_1x_2. \tag{3.25}$$

Finally, the first integrals associated by Noether's theorem to these vector fields are, according to (3.17) and (3.20),

$$\begin{aligned}
&I_1=\mathbf{I}_4^2/2,\quad I_2=\mathbf{I}_4\mathbf{I}_5,\quad I_3=\mathbf{I}_5^2/2,\\
&\mathbf{I}_4=W(t)^{-1}[x_1'(t)\mathbf{x}-x_1(t)\mathbf{x}'-\mathbf{W}_1(t)],\\
&\mathbf{I}_5=W(t)^{-1}[x_2'(t)\mathbf{x}-x_2(t)\mathbf{x}'-\mathbf{W}_2(t)],\\
&I_6^{ij}=I_4^iI_5^j-I_4^jI_5^i,\quad 1\leqslant i<j\leqslant n,
\end{aligned} \tag{3.26}$$

where again $t\in J$. To extend these *local* results, observe that the Lagrangian (3.23) can be written as

$$L_A=L+\lambda', \tag{3.27}$$

where

$$L=\frac{1}{W(t)}\left[\frac{\mathbf{x}'^2}{2}-a_0(t)\frac{\mathbf{x}^2}{2}-\mathbf{b}(t)\mathbf{x}\right], \tag{3.28}$$

and

$$\lambda(t,\mathbf{x})=\frac{1}{2}\int_{t_0}^{t}\frac{\mathbf{W}_1^2(s)}{W(s)x_1^2(s)}ds+\frac{\mathbf{W}_1(t)\mathbf{x}}{x_1(t)W(t)}-\frac{x_1'(t)\mathbf{x}^2}{2x_1(t)W(t)},\quad (t,\mathbf{x})\in V. \tag{3.29}$$

From (3.27) it easily follows that we can replace L_A by L in all the above considerations, since both Lagrangians differ by a total derivative. In other words, the vectors (3.23) and (3.24) are a basis of the variational symmetry algebra of the restriction of (3.1) to V' *relative to L*, and the functions (3.26) are their associated first integrals. Since $W(t)$ is the Wronskian of two linearly independent solutions x_1 and x_2 of the linear ODE (2.36), it never vanishes; hence the Lagrangian (3.28), the vector fields (3.24), and their corresponding first integrals (3.26) are well defined on all of $\mathbf{R}\times\mathbf{R}^n$. By a continuity argument completely analogous to that of Proposition 2 of the last section, we come to the following conclusion.

Theorem 3: A basis of the variational symmetry algebra of the isotropic system (3.1) relative to the Lagrangian (3.28) is provided by the $(n^2+3n+6)/2$ vector fields (3.24). The structure constants of this basis are the same as those of the basis (3.19) of the variational symmetry algebra of $\mathbf{y}''=\mathbf{0}$, and their associated first integrals are the functions (3.26).

Let us now consider an arbitrary, maximally symmetric system (2.1)–(2.47). From Sec. II we know that this system is equivalent to the isotropic system

$$\mathbf{y}''+a(u)\mathbf{y}=\mathbf{0} \tag{3.30}$$

under the global diffeomorphism (2.2). This system is Lagrangian, a Lagrangian being given by

$$L_0=[\mathbf{y}'^2-a(u)\mathbf{y}^2]/2. \tag{3.31}$$

From Theorem 3 and Lemma 1 an additional theorem then follows.

Theorem 4: The maximally symmetric system (2.1)–(2.48) is Lagrangian, a Lagrangian being given by

$$\begin{aligned}L=&\tfrac{1}{2}\{R^{-1}(t)[(\mathbf{x}'-\mathbf{x}_0'(t))+A_1(t)(\mathbf{x}-\mathbf{x}_0(t))]\}^2\\&-\tfrac{1}{2}a(t)[R^{-1}(t)(\mathbf{x}-\mathbf{x}_0(t))]^2,\end{aligned} \tag{3.32}$$

where the matrix $R(t)$ was defined by (1.6). The variational symmetry algebra of (2.1)–(2.47) relative to the Lagrangian (3.32) is isomorphic to the variational symmetry algebra of $\mathbf{y}''=\mathbf{0}$, generated by the $(n^2+3n+6)/2$ vector fields (3.19).

Remark: When (2.1)–(2.47) is isotropic, the Lagrangian (3.32) does not reduce to (3.28), as can easily be checked. However, it is straightforward to check that in this case (3.33) and (3.28) differ just by a total derivative, and therefore Theorem 4 reduces to Theorem 3 for isotropic systems.

We shall finish this section by studying the structure of the Lie algebra $\mathscr{g}^V$ generated by the vector fields (3.19) [or, equivalently, (3.24)], whose commutation relations are listed in Table II. Let us denote by $\mathscr{A}$ the subspace generated by the vectors N_4^i and N_5^i ($1\leqslant i\leqslant 2n$), and let $\mathscr{S}$ be the span of N_1,N_2,N_3, and N_6^{ij} ($1\leqslant i<j\leqslant n$). By construction, $\mathscr{g}^V$ is, as a vector space, the direct sum of its subspaces $\mathscr{A}$ and $\mathscr{S}$. Moreover, from Table I it follows that $\mathscr{A}$ is an ideal of $\mathscr{g}^V$ (i.e., $[\mathscr{A},\mathscr{g}^V]\subset\mathscr{A}$), and $\mathscr{S}$ is a subalgebra (i.e., $[\mathscr{S},\mathscr{S}]\subset\mathscr{S}$). Hence, as a Lie algebra, $\mathscr{g}^V$ is the semidirect sum[23] of $\mathscr{A}$ and $\mathscr{S}$. Let us now investigate the structure of $\mathscr{A}$ and $\mathscr{S}$. First of all, from Table I it follows that $\mathscr{A}$ is an Abelian ideal, i.e., $[\mathscr{A},\mathscr{A}]=\{0\}$. On the other hand, from a cursory inspection of Table I it also follows that the complementary subspaces of $\mathscr{S}$,

$$\mathscr{S}_1=\text{span}\{N_1,N_2,N_3\}, \tag{3.33}$$

$$\mathscr{S}_2=\text{span}\{N_6^{ij}\,|\,1\leqslant i<j\leqslant n\}, \tag{3.34}$$

TABLE II. The variational symmetry algebra.

	N_1	N_2	N_3	N_4^k	N_5^k	N_6^{kl}
N_1	0	$2N_1$	N_2	0	N_4^k	0
N_2	$-2N_1$	0	$2N_3$	$-N_4^k$	N_5^k	0
N_3	$-N_2$	$-2N_3$	0	$-N_5^k$	0	0
N_4^i	0	N_4^i	N_5^i	0	0	$\delta_{ik}N_4^l - \delta_{il}N_4^k$
N_5^i	$-N_4^i$	$-N_5^i$	0	0	0	$\delta_{ik}N_5^l - \delta_{il}N_5^k$
N_6^{ij}	[LP:0:]	[LP:0:]	[LP:0:]	$\delta_{jk}N_4^i - \delta_{ik}N_4^j$	$\delta_{jk}N_5^i - \delta_{ik}N_5^j$	$\delta_{ij}N_6^{il} + \delta_{il}N_6^{jk} + \delta_{jl}N_6^{ki} + \delta_{ik}N_6^{lj}$

are ideals of $\mathscr{S}$, i.e., $[\mathscr{S}_\alpha,\mathscr{S}]\subset\mathscr{S}$ $(\alpha=1,2)$. Therefore $\mathscr{S}$, as a Lie algebra, is the direct sum[18] of these ideals:

$$\mathscr{S}=\mathscr{S}_1\oplus\mathscr{S}_2. \tag{3.35}$$

Finally, it is easily seen that $\mathscr{S}_1$ is isomorphic to sl(2,**R**) and $\mathscr{S}_2$ to so(n,**R**). Indeed, it is straightforward to check that the generators

$$\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix},\quad \begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix},\quad \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix} \tag{3.36}$$

of sl(2,**R**) obey exactly the same algebra as the generators N_1, N_2, and N_3 of $\mathscr{S}_1$. That $\mathscr{S}_2$ is isomorphic to so(n,**R**) is also easy to see, since we know that the vector fields (3.19) satisfy the commutation relations listed in Table I, and in this representation we have

$$\begin{aligned}\mathscr{S}_2&=\operatorname{span}\{V_6^{ij}\,|\,1\leqslant i<j\leqslant n\}\\ &=\operatorname{span}\{y_i\,\partial_j-y_j\,\partial_i\,|\,1\leqslant i<j\leqslant n\},\end{aligned} \tag{3.37}$$

and the vector fields $y_i\,\partial_j-y_j\,\partial_i$ are the well-known generators of the group of proper rotations of $\mathbf{R}^n$. Putting all the above together, we have shown that $\mathscr{G}^V$ has the following structure:

$$\begin{aligned}&\mathscr{G}^V=\mathscr{A}+(\mathscr{S}_1\oplus\mathscr{S}_2),\quad \mathscr{A}\approx\mathbf{R}^{2n},\\ &\mathscr{S}_1\approx \mathrm{sl}(2,\mathbf{R}),\quad \mathscr{S}_2\approx \mathrm{so}(n,\mathbf{R}),\end{aligned} \tag{3.38}$$

where " + " stands for "semidirect sum." To conclude this analysis, we shall show that (3.38) is the Levi–Mal'cev decomposition[18] of $\mathscr{G}^V$. To this end, it suffices to show[23,24] that $\mathscr{S}_1\oplus\mathscr{S}_2$ is semisimple, i.e., that $\mathscr{S}_1$ and $\mathscr{S}_2$ are simple; but this is immediate from the second line of (3.38). Hence (3.38) is indeed the Levi–Mal'cev decomposition of $\mathscr{G}^V$.

The decomposition (3.38) suggests that $\mathscr{G}^V$ is isomorphic to the Lie algebra of the abstract Lie group

$$\mathscr{G}=\mathbf{R}^{2n}\odot(\mathrm{SL}(2,\mathbf{R})\otimes \mathrm{SO}(n,\mathbf{R})), \tag{3.39}$$

where the symbols $\odot$ and $\otimes$ stand for semidirect and direct product,[24] respectively, and the action of SL(2,**R**) $\otimes$ SO(n,**R**) on $\mathbf{R}^{2n}=\mathbf{R}^n\times\mathbf{R}^n$ is the natural one. That this is true can be proved by a straightforward computation. Finally, let us mention that the Lie algebra of (3.39), and hence $\mathscr{G}^V$, admits a simple realization by means of the algebra of $(n+2)\times(n+2)$ matrices of the form

$$\begin{bmatrix}A & 0\\ C & B\end{bmatrix}, \tag{3.40}$$

with $A\in\mathrm{SL}(2,\mathbf{R})$, $B\in\mathrm{SO}(n,\mathbf{R})$, and C an arbitrary $n\times 2$ matrix.

IV. EXAMPLES

In this section we shall apply Eqs. (2.43) and (3.24) to computing the symmetry algebra and the variational symmetry algebra of the harmonic oscillator with time-dependent frequency in n dimensions:

$$\mathbf{x}''+a(t)\mathbf{x}=\mathbf{0}\quad(\mathbf{x}\in\mathbf{R}^n). \tag{4.1}$$

[Physically, $a(t)$ must be positive for (4.1) to represent a harmonic oscillator of frequency $\omega(t)=[a(t)]^{1/2}$, but this restriction shall be unnecessary in what follows.] As is customary,[10,11,15] we shall choose a basis of solutions $x_1(t)$, $x_2(t)$ of the associated scalar ODE

$$x''+a(t)x=0\quad(x\in\mathbf{R}), \tag{4.2}$$

of the form

$$x_1=\rho\cos\theta,\quad x_2=\rho\sin\theta, \tag{4.3}$$

where $\rho(t)>0$ is a particular solution of Pinney's (nonlinear) equation,[25]

$$\rho''+a(t)\rho=\rho^{-3}\quad(\rho>0), \tag{4.4}$$

and $\theta(t)$ is defined in terms of $\rho(t)$ as follows:

$$\theta(t)=\int_{t_0}^{t}\frac{ds}{\rho^2(s)}. \tag{4.5}$$

[The solutions of (4.4) are well defined for all t; indeed, a straightforward calculation shows that the solution of (4.3) with initial conditions $\rho(0)=\rho_0$, $\rho'(0)=\rho_0'$ can be written as follows:

$$\rho=[(\rho_0 y_1(t)+\rho_0'\,y_2(t))^2+\rho_0^{-2}y_2^2(t)]^{1/2}, \tag{4.6}$$

with y_1 and y_2 a fundamental basis of solutions of (4.2).] The Wronskian of (4.3) is easily computed, obtaining

$$W(t)=1,\quad \text{all } t\in\mathbf{R}. \tag{4.7}$$

Moreover, we obviously have

$$\mathbf{x}_0=\mathbf{W}_1=\mathbf{W}_2=\mathbf{0}, \tag{4.8}$$

since (4.1) is homogeneous. Inserting (4.3), (4.7), and (4.8) into (2.43), we obtain the following basis of the symmetry algebra of (4.1):

$$\begin{aligned}&X_1=\rho\cos\theta\,[\,\rho\cos\theta\partial_t+(\rho'\cos\theta-\rho^{-1}\sin\theta)\mathbf{x}\,\partial_{\mathbf{x}}\,],\\ &X_2=\rho\sin\theta\,[\,\rho\cos\theta\partial_t+(\rho'\cos\theta-\rho^{-1}\sin\theta)\mathbf{x}\,\partial_{\mathbf{x}}\,]\\ &X_3^i=x_i[\,\rho\cos\theta\,\partial_t+(\rho'\cos\theta-\rho^{-1}\sin\theta)\mathbf{x}\,\partial_{\mathbf{x}}\,],\\ &X_4^i=\rho\cos\theta\,\partial_i,\quad X_5^i=\rho\sin\theta\,\partial_i,\quad X_6^{ij}=x_i\,\partial_j,\\ &X_7=\rho\sin\theta\,[\,\rho\sin\theta\,\partial_t+(\rho'\sin\theta+\rho^{-1}\cos\theta)\mathbf{x}\,\partial_{\mathbf{x}}\,],\end{aligned} \tag{4.9}$$

$$X_8^i = x_i[\rho\cos\theta\partial_t + (\rho'\cos\theta - \rho^{-1}\sin\theta)\mathbf{x}\,\partial_\mathbf{x}],$$
$$1\leqslant i,j\leqslant n.$$

By Proposition 2 of Sec. II we know, without having to perform any calculation, that these vector fields satisfy the commutation relations listed in Table I. Defining

$$\begin{aligned} &G_1 = 2X_2 + \sum_{1\leqslant i\leqslant n} X_6^{ii}, \quad G_2 = X_1 - X_7, \quad G_3^i = X_4^i, \\ &G_4^i = X_5^i, \quad G_5 = X_1 + X_7, \quad G_6^{ij} = X_6^{ij}, \\ &G_7^i = X_8^i, \quad G_8^i = X_3^i, \quad 1\leqslant i,j\leqslant n, \end{aligned} \tag{4.10}$$

we obtain the generators found in Ref. 7.

According to (3.28), a Lagrangian for the system (4.1) is

$$L = [\mathbf{x}'^2 - a_0(t)\mathbf{x}^2]/2. \tag{4.11}$$

The generators of the symmetry algebra of (4.1) with respect to this Lagrangian can be easily computed using (3.23) or (3.24), yielding

$$\begin{aligned} &N_1 = \rho\cos\theta\,[\rho\cos\theta\,\partial_t + (\rho'\cos\theta - \rho^{-1}\sin\theta)\mathbf{x}\,\partial_\mathbf{x}], \\ &N_2 = \rho^2\sin 2\theta\,\partial_t + (\rho\rho'\sin 2\theta + \cos 2\theta)\mathbf{x}\,\partial_\mathbf{x}, \\ &N_3 = \rho\sin\theta\,[\rho\sin\theta\,\partial_t + (\rho'\sin\theta + \rho^{-1}\cos\theta)\mathbf{x}\,\partial_\mathbf{x}], \\ &N_4^i = \rho\cos\theta\,\partial_i, \quad N_5^i = \rho\sin\theta\,\partial_i, \\ &N_6^{ij} = x_i\partial_j - x_j\partial_i, \quad 1\leqslant i<j\leqslant n. \end{aligned} \tag{4.12}$$

Using (3.26) we obtain the first integrals associated to these generators by Noether's theorem:

$$\begin{aligned} I_1 &= [\rho^2\cos^2\theta\,\mathbf{X}'^2 + (\sin 2\theta - 2\rho\rho'\cos^2\theta)\mathbf{x}\mathbf{x}' \\ &\quad + (\rho'\cos\theta - \rho^{-1}\sin\theta)\mathbf{x}^2]/2, \\ I_2 &= \{\rho^2\sin^2\theta\mathbf{x}'^2 - 2(\rho\rho'\sin 2\theta + \cos 2\theta)\mathbf{x}\mathbf{x}' \\ &\quad + [2\rho^{-1}\rho'\cos\theta + (\rho'^2 - \rho^{-2})\sin 2\theta]\mathbf{x}^2\}/2, \\ I_3 &= [\rho^2\sin^2\theta\mathbf{x}'^2 - (\sin 2\theta + 2\rho\rho'\cos^2\theta)\mathbf{x}\mathbf{x}' \\ &\quad + (\rho'\sin\theta + \rho^{-1}\cos\theta)\mathbf{x}^2]/2, \\ \mathbf{I}_4 &= -\rho\cos\theta\mathbf{x}' + (\rho'\cos\theta - \rho^{-1}\sin\theta)\mathbf{x}, \\ \mathbf{I}_5 &= -\rho\sin\theta\mathbf{x}' + (\rho'\sin\theta + \rho^{-1}\cos\theta)\mathbf{x}, \\ I_6^{ij} &= x_i x_j' - x_j x_i', \quad 1\leqslant i<j\leqslant n. \end{aligned} \tag{4.13}$$

Replacing the generators N_1 and N_3 in (4.12) by their linear combinations $N_1 \pm N_3$, we obtain the basis found in Ref. 7. Again, by Theorem 1 of Sec. III the generators (4.12) satisfy the commutation relations listed in Table II.

ACKNOWLEDGMENTS

It is a pleasant duty to express our gratitude to Professor F. González Gascón for his valuable suggestions and encouragement. We would also like to thank the Physics Department of Princeton University, where this work was completed, for its warm hospitality.

Finally, we gratefully acknowledge the financial support of Fulbright–MEC Grant No. 86-07391.

[1]F. González Gascón, Phys. Lett. A **61**, 375 (1977); J. Math. Phys. **21**, 2046 (1980); Lett. Nuovo Cimento **34**, 35, 225 (1982).

[2]F. González Gascón and A. González López, Lett. Nuovo Cimento **39**, 315 (1984).

[3]F. González Gascón, Lett. Nuovo Cimento **22**, 59 (1978); **29**, 73 (1980).

[4]C. E. Wulfman and B. G. Wybourne, J. Phys. A **9**, 507 (1976).

[5]L. Dolan, Phys. Rep. C **109**, 1 (1984), and references therein.

[6]P. J. Olver, *Applications of Lie Groups to Differential Equations* (Springer, New York, 1986).

[7]G. E. Prince and C. J. Eliezer, J. Phys. A **14**, 587 (1981).

[8]F. González Gascón and A. González López, J. Math. Phys. **24**, 2006 (1983).

[9]S. Lie, *Differentialgleichungen* (Chelsea, New York, 1967), reprint.

[10]G. E. Prince and C. J. Eliezer, J. Phys. A **13**, 815 (1980). See also R. L. Anderson and S. M. Davison, J. Math. Anal. Appl. **48**, 301 (1974); C. E. Wulfman and B. G. Wybourne, J. Phys. A **9**, 507 (1976); and M. Lutzky, J. Phys. A **11**, 249 (1978), for the one-dimensional oscillator; and P. G. L. Leach, J. Math. Phys. **21**, 300 (1980); J. Phys. A **21**, 1991 (1980), for the one-dimensional oscillator with time-dependent frequency and the one-dimensional "repulsive" oscillator, respectively.

[11]J. M. Cerveró and J. Villarroel, J. Phys. A **17**, 1777 (1984).

[12]P. G. L. Leach, "Symmetries and transformations for one-dimensional classical dynamical systems (1979)," in *Group Theory, O.D.E.'s and Classical Mechanics* (La Trobe University, Bundoora, Australia, 1982).

[13]R. Martini and P. H. M. Kersten, J. Phys. A **16**, L455 (1983).

[14]M. Aguirre and J. Krause, J. Math. Phys. **26**, 593 (1985).

[15]P. G. L. Leach, J. Phys. A **13**, 1991 (1980).

[16]S. Lie, *Theorie der Transformationengruppe I, II, III* (Teubner, Leipzig, 1888, 1890, 1893).

[17]S. Helgason, Bull. Am. Math. Soc. **83**, 751 (1977), and references therein.

[18]N. Jacobson, *Lie Algebras* (Interscience, New York, 1962).

[19]V. Arnold, *Geometric Methods in the Theory of Ordinary Differential Equations* (Springer, New York, 1983).

[20]Y. Matsushima, *Differential Manifolds* (Marcel Dekker, New York, 1972), p. 85.

[21]P. Havas, Nuovo Cimento Suppl. V, Ser. X, 363 (1957).

[22]W. Sarlet and F. Cantrijn, SIAM Rev. **23**, 467 (1981).

[23]N. H. Ibragimov, *Transformation Groups Applied to Mathematical Physics* (Reidel, Dordrecht, Holland, 1985), §2.

[24]L. Jansen and M. H. Boon, *Group Theory* (North-Holland, Amsterdam, 1967), Chap. 1, §6.

[25]E. Pinney, Proc. Am. Math. Soc. **1**, 681 (1950).

On the denominator function for canonical SU(3) tensor operators. II. Explicit polynomial form

J. D. Louck and L. C. Biedenharn
Los Alamos National Laboratory, Theoretical Division, Los Alamos, New Mexico 87545
M. A. Lohe
University College of the Northern Territory, Darwin NT 5794, Australia

(Received 30 November 1987; accepted for publication 30 December 1987)

The canonical resolution of the multiplicity problem for tensor operators in SU(3) is equivalent to the map (the denominator mapping) from the set of all SU(3) unit tensor operators to SU(3) invariant functions (the denominator functions). The denominator function vanishes precisely on that characteristic null space that specifies each operator uniquely since [for SU(3)] the characteristic null spaces are known to be simply ordered. Each denominator function can be expressed, up to explicitly known multiplicative factors, as a ratio of two successive polynomials in the set $\{G_q^t\}$, $t = 0,1,..., q+1$, $q = 0,1,...$. By obtaining explicitly the set of all polynomials $\{G_q^t\}$, this paper completes the construction of all SU(3) denominator functions.

I. INTRODUCTION

In a previous paper[1] (referred to here as I), we have defined the denominator function for SU(3) tensor operators and determined its symmetry properties and reduction formulas. The significance of the denominator function—denoted by $D^2(\Gamma_t,x)$, see Eq. (1.3)—is that this function defines uniquely (to within sign conventions) the canonical construction of the unit tensor operators of the symmetry group SU(3). [This canonical construction (in terms of the characteristic null spaces of the operators) is discussed in detail in I.] Explicit (algebraic) construction of the denominator function is a basic and major step toward determining full algebraic expressions for SU(3) Wigner–Clebsch–Gordan (WCG) coefficients (matrix elements of canonical unit tensor operators).

The initial form obtained in I for the denominator function—as implied by the canonical splitting of the multiplicity[2-6]—is extremely complicated and unwieldy; $D^2(\Gamma_t,x)$ appears as the ratio of two determinants whose elements are themselves complicated functions [see Eqs. (2.23)–(2.25) in I]. Considerable simplification results upon recognizing that the denominator function can be expressed, to within multiplicative factors, as a ratio of polynomials from the set denoted $\{G_q^t\}$. Details of this construction were given in I, but expressions for the G_q^t as polynomials were not obtained there.

In the present paper, we complete the development and verify the properties of the polynomials $G_q^t(\Delta;x)$. It is our goal to prove that

$$G_q^t(\Delta;x) = \mathscr{G}_q^t(\Delta;x), \tag{1.1}$$

where $\mathscr{G}_q^t(\Delta;x)$ is an explicitly defined polynomial of total degree $2t(q-t+1)$ both in the barycentric coordinates $x = (x_1,x_2,x_3)$ and the shift coordinates $(\Delta_1,\Delta_2,\Delta_3)$ [see Eqs. (4.2)]. We are able to reduce the proof to one of showing that the polynomial $\mathscr{G}_q^t(\Delta;x)$ satisfies (identically in all variables) the relation

$$\mathscr{G}_q^t(\Delta_1,\Delta_2,\Delta_3;\ x_1,x_2,x_3) = \mathscr{G}_q^t(\Delta_2 - x_3,\ \Delta_1 + x_3,\ \Delta_3;\ x_1,x_2,x_3). \tag{1.2}$$

In other words, the validity of relation (1.2) implies that of (1.1).

In principle, the proof of relation (1.2) should involve only straightforward verification since $\{\mathscr{G}_q^t(\Delta;x)\}$ is a set of explicitly given polynomials. Despite the elegance and apparent simplicity of the polynomials $\{\mathscr{G}_q^t(\Delta;x)\}$, the proof of property (1.2) turns out to be unexpectedly difficult. We have indeed constructed a proof, but the method is so foreign to those of this paper as to require separate publication.[7] (In a sense, one can even avoid the problem by properly symmetrizing $\mathscr{G}_q^t$, but this is inelegant and unsatisfactory.)

The strategy of the present paper for proving relation (1.1) is to establish a (nonconstructive) uniqueness theorem. We show that the following three properties uniquely determine, up to a multiplicative function of $\Delta_1+\Delta_2+\Delta_3$, a polynomial $Q_q^t(\Delta;x)$: (i) total degree $2t(q-t+1)$ in x, (ii) determinantal symmetry, and (iii) the weight space set of zeros. [Properties (ii) and (iii) are explained more fully below (see, also, I).]

Properties (i)–(iii) of $G_q^t(\Delta;x)$ were proved in I, as reviewed below. Properties (i), (iii), and part of (ii) of $\mathscr{G}_q^t(\Delta;x)$ are proved here; the proof of relation (1.2) then establishes all of property (ii) and, together with the uniqueness theorem, will be used to prove the desired result, Eq. (1.1).

The paper is organized as follows: In the present section, we give the relation between the denominator function $D^2(\Gamma_t;x)$ and the explicitly defined function $G_q^t(\Delta;x)$, whose main properties (proved in I) are also summarized. In Sec. II, we prove two lemmas for any polynomial having the weight space set of zeros. In Sec. III, we prove the uniqueness theorem mentioned above. In Sec. IV, we give the explicit polynomial $\mathscr{G}_q^t(\Delta;x)$ and develop some of its properties. In particular, we prove that $\mathscr{G}_q^t$ satisfies (i)–(iii)

 0022-2488/88/051106-12$02.50

above, provided relation (1.2) is true. This then allows us to prove Eq. (1.1). In Sec. V, we give a necessary condition for relation (1.2) to be true. It is the proof of this latter relation that will be given in Ref. 7.

Before summarizing the properties of the polynomials $G_q^t(\Delta;x)$, we first define some symbols that are used throughout.

(i) The set of real numbers and the set of non-negative integers are denoted by $\mathbb{R}$ and $\mathbb{Z}_0$, respectively.

(ii) The Möbius plane and the subset of (lattice) points of $\mathbb{M}$ with integral coordinates are denoted by $\mathbb{M}$ and $\mathbb{L}$, respectively.

(iii) Integers such that $q\in\mathbb{Z}_0, p=q,q+1,\ldots$ are denoted by q and p.

(iv) The three-tuple $(\Delta_1,\Delta_2,\Delta_3)$ such that $\Delta_i\in\mathbb{Z}_0$, $0\leqslant\Delta_i\leqslant p$, and $\Delta_1+\Delta_2+\Delta_3=p+q$ is denoted $\Delta=(\Delta_1,\Delta_2,\Delta_3)$. (These conditions are sometimes relaxed to $\Delta\in\mathbb{R}^3$, but this will be clear from the context.)

(v) A point in $\mathbb{M}$, which is sometimes restricted to $\mathbb{L}$, is denoted $x=(x_1,x_2,x_3)$.

(vi) Pochhammer's notation for the rising factorial for $a\in\mathbb{Z}_0$ with $(x)_0=1$ is $(x)_a=x(x+1)\cdots(x+a-1)$; similarly, $[x]_a=x(x-1)\cdots(x-a+1)$ denotes a falling factorial.

The basic problem addressed in this paper is to prove the equality of two functions, $G_q^t(\Delta;x)$ and $\mathscr{G}_q^t(\Delta;x)$, each of which is known. Thus the problem can be stated quite independently of its origin in the theory of SU(3) tensor operators. In the interest of clarity and accessibility, we give both of these functions [Eqs. (1.4) and (4.2)], noting only briefly the relationship to the denominator function that occurs in the unit tensor operator coefficients of SU(3). The goal then is the proof of the identity (1.1).

The SU(3) denominator function $D^2(\Gamma_t;x)$ is defined in terms of the function $G_q^t(\Delta;x)$ by

$$\frac{1}{D^2(\Gamma_t;x)}=\frac{(-1)^{\Delta_2-t+1}}{C_{p,q}^t}\frac{\mathrm{Dim}(x^f)}{\mathrm{Dim}(x)}\times\frac{1}{L_t(\Delta;x)}\frac{G_q^t(\Delta;x)}{G_q^{t-1}(\Delta;x)},\tag{1.3a}$$

where

$$L_t(\Delta;X)=\prod_{ijk}(\Delta_i-t+1)!(x_i+1)_{\Delta_j-t+1}\times(-x_i+1)_{\Delta_k-t+1}\quad(ijk\text{ cyclic}),\tag{1.3b}$$

$$C_{p,q}^t=\frac{(t-1)!(p-t+2)!(q-t+1)!}{(p-q)!(p+1)!q!},\tag{1.3c}$$

$$\mathrm{Dim}(x)=-x_1x_2x_3/2,\tag{1.3d}$$

$$\mathrm{Dim}(x^f)=-x_1^fx_2^fx_3^f/2,$$
$$x_i^f=x_i+\Delta_j-\Delta_k\quad(ijk\text{ cyclic}).\tag{1.3e}$$

Here $t=1,2,\ldots,\mathscr{M}$ where $\mathscr{M}$ denotes the multiplicity of weight Δ in the irrep $[p,q,0]$ of SU (3). For each value of t the corresponding denominator function $D^2(\Gamma_t;x)$ belongs to that canonical unit tensor operator which is characterized by the operator pattern Γ_t (see I).

The function $G_q^t(\Delta;x)$ in definition (1.3a) has the following complicated form, as derived directly from the canonical construction of the set of SU(3) tensor operators:

$$G_q^t(\Delta;x)=(-1)^{t(q+1)}\prod_{s=1}^{t}\frac{(s-1)!(p-s+2)!(q-s+1)!}{(p+1)!}\times\prod_{s=1}^{t}\frac{1}{\Pi_{ijk}(-\Delta_i)_{s-1}(-x_i-\Delta_j)_{s-1}(x_i-\Delta_k)_{s-1}}\times\left[\frac{(q!)^3}{\mathrm{Dim}(x)L_1(qqq;x)}\right]^t\bar{A}_t(\Delta,\lambda;x).\tag{1.4a}$$

The quantities appearing in expression (1.4a) have the following definitions: $\bar{A}_t(\Delta,\lambda;x)$ is the $t\times t$ determinant defined by

$$\bar{A}_t(\Delta,\lambda;x)=\det(N^{rs}),\tag{1.4b}$$

with entry in row r and column s given by

$$N^{rs}=N^{rs}(\Delta,\lambda;x)=\sum_{n\in\mathbb{D}}F_n^r(\lambda;x)F_n^s(\lambda;x)N_n(\Delta;x).\tag{1.4c}$$

In relation (1.4c), the $N_n(\Delta;x)$ are the functions defined by

$$N_n(\Delta;x)=\frac{(x_1-n_2+n_3)(x_2-n_3+n_1)(x_3-n_1+n_2)}{2n_1!n_2!n_3!}\times\prod_{ijk}(-\Delta_i)_{q-n_i}(x_i-\Delta_k)_{q-n_j}\times(-x_i-\Delta_j)_{q-n_k}(x_i-q)_{q-n_j}(-x_i-q)_{q-n_k},\quad\text{for each } n\in\mathbb{D},\tag{1.4d}$$

and the $F_n^t(\lambda;x)$ are the functions defined by the expansion

$$\prod_{ijk}\left(y+1+\frac{1}{3}(x_j-x_k)+\lambda_i\right)_{n_i-\lambda_i}=\sum_{t=1}^{\mathscr{M}}y^{\mathscr{M}-t}F_n^t(\lambda;x),\tag{1.4e}$$

$$\lambda_i=\max(0,q-\Delta_i),\tag{1.4f}$$

$$\mathscr{M}=q+1-\lambda_1-\lambda_2-\lambda_3,\tag{1.4g}$$

$$\prod_{ijk}A_{ijk}=A_{123}A_{231}A_{312}.\tag{1.4h}$$

Finally, n is the three-tuple of integers (n_1,n_2,n_3), with domain $\mathbb{D}$ defined for specified p,q,Δ by

$$\mathbb{D}=\mathbb{D}(p,q,\Delta)=\left\{(n_1,n_2,n_3)\,\middle|\,\begin{array}{l}n_i\in\{\lambda_i,\lambda_i+1,\ldots,\sigma_i\}\\ n_1+n_2+n_3=q\end{array}\right\},\tag{1.4i}$$

where σ_i is defined by $\sigma_i=\min(q,p-\Delta_i)$. The quantities $\mathrm{Dim}(x)$ and $L_1(qqq;x)$ are those defined by Eqs. (1.3d) and (1.3b), respectively, with the latter being for the special case $t=1$ and $\Delta_1=\Delta_2=\Delta_3=q$.

As noted above, the form of $G_q^t(\Delta;x)$ given by Eqs. (1.4) is indeed very complicated. Despite this, it has been possible in I to prove many significant properties of $G_q^t(\Delta;x)$. These properties are, in fact, definitive in establishing a comprehensible (new) expression for this function, which is the purpose of the present paper.

Let us now summarize from I the properties of the function $G_q^t(\Delta;x)$ that will be used below for obtaining an alternative and explicit form. For each $q=0,1,2,\ldots$ and $t=0,1,\ldots,q+1$, the function $G_q^t(\Delta;x)$ is a polynomial that has the following properties.

(i) Total degree $2t(q-t+1)$ in x. By this phrase, we mean that $G_q^t(\Delta;x)$ is a sum of monomials of the form $x_1^\alpha x_2^\beta x_3^\gamma$, where α, β, γ are non-negative integers such that $\alpha+\beta+\gamma \leqslant 2t(q-t+1)$ and the sum is over all such monomials multiplied by real coefficients that are themselves functions of Δ. For $t=0$ and arbitrary t, we have, by definition,

$$G_t^0(\Delta;x) = G_t^{t+1}(\Delta;x) = 1. \tag{1.5}$$

(ii) Determinantal symmetry. This symmetry refers to the invariance of $G_q^t(\Delta;x)$ under the transformation of the six variables $(\Delta_1,\Delta_2,\Delta_3,x_1,x_2,x_3)$ induced by row interchange, column interchange, and transposition of the 3×3 array A defined by

$$\begin{aligned} A &= A_t(\Delta;x) \\ &= \begin{bmatrix} \Delta_1-t+1 & \Delta_2-t+1+x_1 & \Delta_3-t+1-x_1 \\ \Delta_2-t+1 & \Delta_3-t+1+x_2 & \Delta_1-t+1-x_2 \\ \Delta_3-t+1 & \Delta_1-t+1+x_3 & \Delta_2-t+1-x_3 \end{bmatrix} \\ &= (a_{ij}). \end{aligned} \tag{1.6}$$

For example, under matrix transposition of A, that is, $A \to \widetilde{A}$, we have

$$\begin{aligned} &(\Delta_1,\Delta_2,\Delta_3,x_1,x_2,x_3) \\ &\quad \to (\Delta_1,\Delta_2+x_1,\Delta_3-x_1,-x_1,-x_3,-x_2). \end{aligned} \tag{1.7}$$

(iii) Weight space $\mathbb{W}_q^t(\Delta)$ of zeros. This subset of $\mathbb{L}$ is shown in Fig. 1. The points in $\mathbb{W}_q^t(\Delta)$ are in one-to-one correspondence with those of the weight space of irrep $[q-t,0,-t+1]$ of U(3). With each point $x\in\mathbb{W}_q^t(\Delta)$, we associate a multiplicity number $M_q^t(\Delta;x)$,

$$M_q^t(\Delta;x) \equiv \min\Big(t, q-t+1, 1+d_t(x)\Big), \tag{1.8}$$

where $d_t(x)$ is the "distance" from lattice point $x\in\mathbb{W}_q^t(\Delta)$ to the nearest boundary point as measured along the direction of a coordinate axis (one lattice spacing = one unit of distance, with $d_t=0$ at the boundary). The multiplicity function $M_q^t(\Delta;x)$ assigns to each point $x\in\mathbb{W}_q^t(\Delta)$ exactly the value of the multiplicity of the weight $w=(w_1,w_2,w_3)$ of irrep $[q-t,\ 0,\ -t+1]$, where w is related to the point $x\in\mathbb{W}_q^t(\Delta)$ by

$$\begin{aligned} x_1 &= \Delta_3 - t + 1 - w_1, \\ x_2 &= -\Delta_2 - \Delta_3 + q - 1 - w_2, \\ x_3 &= \Delta_2 - t + 1 - w_3. \end{aligned} \tag{1.9}$$

By the phrase "a polynomial has the weight space $\mathbb{W}_q^t(\Delta)$ of zeros," we mean that each $x\in\mathbb{W}_q^t(\Delta)$ is a zero of the polynomial with multiplicity $M_q^t(\Delta;x)$.

(iv) Reduction formula. This property refers to an identity satisfied by $G_q^t(\Delta;x)$ when the Δ_i are restricted in their values [see Eq. (4.15) of I]. In particular, for $t-1 \leqslant \Delta_1 \leqslant q$, $\Delta_2\in\mathbb{Z}_0$, $\Delta_3\in\mathbb{Z}_0$, the following identity is true:

$$\begin{aligned} &G_q^t(\Delta_1,\Delta_2,\Delta_3;x_1,x_2,x_3) \\ &\quad = (-1)^{t(q-\Delta_1)} \prod_{s=1}^{t} (-\Delta_2+s-1)_{q-\Delta_1} \\ &\qquad \times (-\Delta_3+s-1)_{q-\Delta_1} (x_1-\Delta_3+s-1)_{q-\Delta_1} \\ &\qquad \times (-x_1-\Delta_2+s-1)_{q-\Delta_1} G_{\Delta_1}^t(q,\Delta_2+\Delta_1-q, \\ &\qquad \Delta_3+\Delta_1-q; x_1, x_2-\Delta_1+q, x_3+\Delta_1-q). \end{aligned} \tag{1.10}$$

The reduction formula (1.10) and the (proved) determinantal symmetry of $G_q^t(\Delta;x)$ may be used to derive a more general reduction formula. To obtain this result, we notice that the arrays A corresponding to the arguments of G_q^t on the left-hand side (lhs) of Eq. (1.10) may be written as

$$A_{11}(h) = (A)_{a_{11}=h-t}, \quad h=t,t+1,\ldots,q+1. \tag{1.11a}$$

Similarly, the arrays A corresponding to the arguments of $G_{\Delta_1}^t$ on the right-hand side (rhs) of Eq. (1.10) are given by

$$A'_{11}(h) = A_{11}(h) + S_{11}(h), \; h=t,t+1,\ldots,q+1, \tag{1.11b}$$

where the "shift matrix" $S_{11}(h)$ is defined by

$$S_{11}(h) = (q-h+1)\begin{bmatrix} 1 & -1 & -1 \\ -1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}. \tag{1.11c}$$

The multiplicative factors in Eq. (1.10) may also be written as

$$\begin{aligned} &(-1)^{t(q-h+1)} \prod_{s=1}^{t} (-a_{12}-s+1)_{q-h+1} \\ &\quad \times (-a_{13}-s+1)_{q-h+1} (-a_{21}-s+1)_{q-h+1} \\ &\quad \times (-a_{31}-s+1)_{q-h+1}. \end{aligned} \tag{1.11d}$$

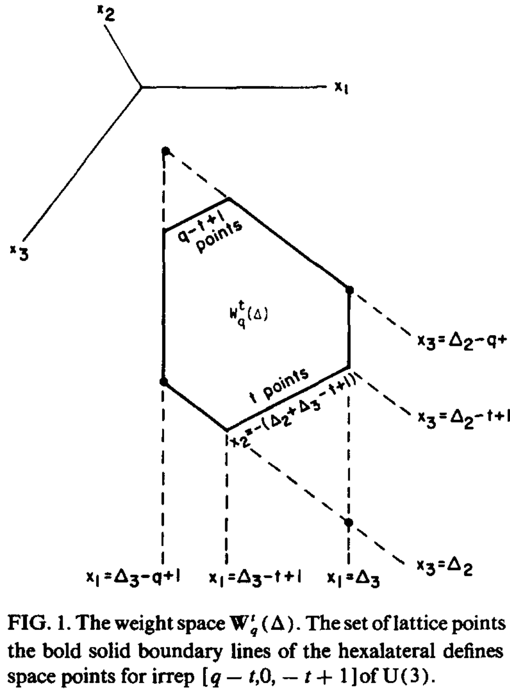


FIG. 1. The weight space $\mathbb{W}_q^t(\Delta)$. The set of lattice points interior to and on the bold solid boundary lines of the hexalateral defines the set of weight space points for irrep $[q-t,0,-t+1]$ of U(3).

Equations (1.11) and determinantal symmetry now imply the following *general reduction formula*:

$$G^t_q(A_{ij}(h))$$
$$= (-1)^{t(q-h+1)} \prod_{s=1}^{t}\left(\prod_{\substack{k=1\\k\neq j}}^{3}(-a_{ik}-s+1)_{q-h+1}\right.$$
$$\left.\times \prod_{\substack{k=1\\k\neq i}}^{3}(-a_{kj}-s+1)_{q-h+1}\right)$$
$$\times G^t_{h-1}(A'_{ij}(h)), \tag{1.12a}$$

for each $h = t, t+1,...,q+1$, where

$$A_{ij}(h) = (A)_{a_{ij}=h-t}, \tag{1.12b}$$
$$A'_{ij}(h) = A_{ij}(h) + S_{ij}(h). \tag{1.12c}$$

Here the shift matrix $S_{ij}(h)$ is the 3×3 array having $q-h+1$ in row i and column j, zeros in the (i,j) minor, and $-(q-h+1)$ elsewhere [see Eq. (1.11c) for $(i,j)=(1,1)$].

(v) Explicit polynomial forms. Since $G^t_{t-1}=1$, we obtain the following fully explicit special cases of the polynomials G^t_q for $h=t$:

$$G^t_q(A)\,|_{a_{ij}=0}$$
$$= (-1)^{t(q-t+1)} \prod_{s=1}^{t}\left(\prod_{\substack{k=1\\k\neq j}}^{3}(-a_{ik}-s+1)_{q-t+1}\right.$$
$$\left.\times \prod_{\substack{k=1\\k\neq i}}^{3}(-a_{kj}-s+1)_{q-t+1}\right). \tag{1.13}$$

This completes our review of the most significant properties of G^t_q proved in I. A principal result of the present paper is the proof that the conjectured form of G^t_q—that is, the polynomial $\mathscr{G}^t_q$ given in Sec. IV—also obeys Eqs. (1.12) and (1.13).

II. GENERAL PROPERTIES OF POLYNOMIALS WITH WEIGHT SPACE $W^t_q(\Delta)$ OF ZEROS

Relation (1.8) may be used to derive explicit properties of an arbitrary polynomial in (x_1,x_2,x_3) that has the weight space $W^t_q(\Delta)$ of zeros. We now use this relation to prove the following lemma.

Lemma 2.1: Each polynomial $P^t_q(\Delta;x)$ having at least the weight space $W^t_q(\Delta;x)$ of zeros contains, for $x_3=\Delta_2-h+1$, the factors

$$\prod_{s=1}^{h}(x_1-\Delta_3+t-s)_{q-t+1}, \quad \text{for } h\in\{1,2,...,t\} \tag{2.1a}$$

and

$$\prod_{s=1}^{t}(x_1-\Delta_3+s-1)_{q-h+1}, \quad \text{for } h\in\{t,t+1,...,q\}. \tag{2.1b}$$

Conversely, each polynomial $P^t_q(\Delta;x)$ that contains the factor (2.1a) for each $h\in\{1,2,...,t\}$ and the factor (2.1b) for each $h\in\{t,t+1,...,q\}$, when evaluated at $x_3=\Delta_2-h+1$, has at least the weight space $W^t_q(\Delta)$ of zeros.

Proof: Let us note first that the lines $x_3=\Delta_2-h+1$ for $h=1,2,...,q$ cover the set of points in $W^t_q(\Delta)$ [see Fig. 2(a)]. Consider now the lines $x_3=\Delta_2-h+1$ for $h\in\{1,2,...,t\}$ as shown in Fig. 2(a) and take $t\leqslant q-t+1$. Then, from Eq. (1.8), we have

$$M^t_q(\Delta;x) = 1 + d_t(x). \tag{2.2}$$

From Fig. 2(a), we find that for $x_1=\Delta_3-q+1$, $\Delta_3-q+2,...,\Delta_3-q+h$ (the x_1 line through point P_1), the value of $1+d_t(x)$ is $1,2,...,h$, respectively; for $x_1=\Delta_3-q+h,\Delta_3-q+h+1,...,\Delta_3-t+1$ (the x_1 line through point P_2), the value of $1+d_t(x)$ is h; and for $x_1=\Delta_3-t+1,\Delta_3-t+2,...,\Delta_3-t+h$ (the x_1 line

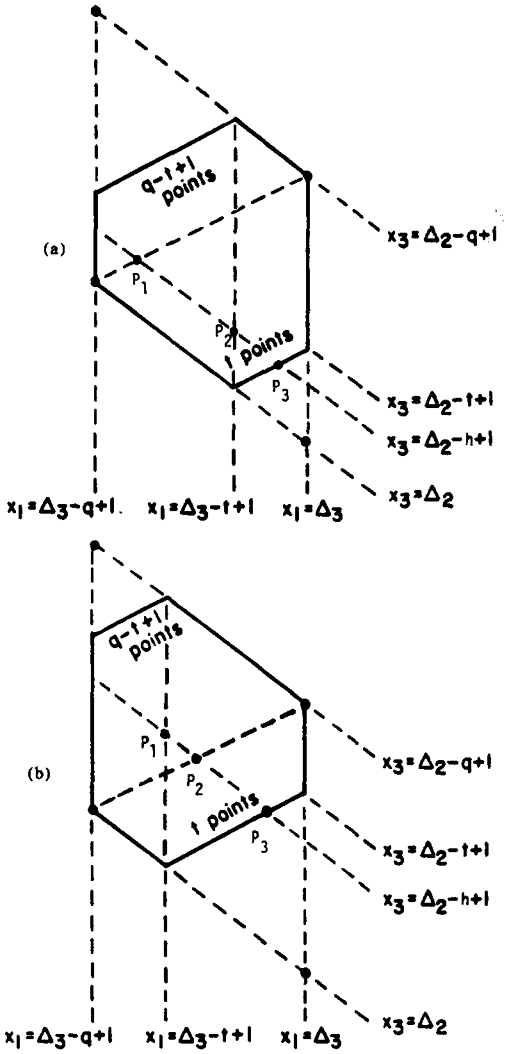


FIG. 2. (a) Covering of the weight space $W^t_q(\Delta)$ for $t\leqslant q-t+1$ by lattice points on the lines $x_3=\Delta_2-h+1$, $h=1,2,...,q$. The multiplicity formula (2.2) may be applied to each point (x_1,x_2,Δ_2-h+1) belonging to $W^t_q(\Delta)$ and to the line $x_3=\Delta_2-h+1$ to obtain the factors given in Lemma 2.1 of an arbitrary polynomial having at least the weight space $W^t_q(\Delta)$ of zeros. (b) Covering of the weight space $W^t_q(\Delta)$ for $t\geqslant q-t+1$ by lattice points on the lines $x_3=\Delta_2-h+1$, $h=1,2,...,q$. The multiplicity formula (2.2) may be applied to each point (x_1,x_2,Δ_2-h+1) belonging to $W^t_q(\Delta)$ and to the line $x_3=\Delta_2-h+1$ to obtain the factors given in Lemma 2.1 of an arbitrary polynomial having at least the weight space $W^t_q(\Delta)$ of zeros.

through point P_3), the value of $1+d_t(x)$ is $h,h-1,\ldots,1$, respectively. Thus, for each $x_3=\Delta_2-h+1$ with $h=1,2,\ldots,t$, the polynomial $P_q^t(\Delta;x)$ must contain factors

$$\begin{aligned}
&(x_1-\Delta_3+q-1)(x_1-\Delta_3+q-2)^2\\
&\quad\times\cdots(x_1-\Delta_3+q-h+1)^{h-1}(x_1-\Delta_3+q-h)^h\\
&\quad\times(x_1-\Delta_3+q-h-1)^h\cdots(x_1-\Delta_3+t-1)^h\\
&\quad\times(x_1-\Delta_3+t-2)^{h-1}(x_1-\Delta_3+t-3)^{h-2}\\
&\quad\times\cdots(x_1-\Delta_3+t-h)\\
&\quad=(x_1-\Delta_3+t-1)_{q-t+1}(x_1-\Delta_3+t-2)_{q-t+1}\\
&\qquad\times\cdots(x_1-\Delta_3+t-h)_{q-t+1},
\end{aligned}\tag{2.3}$$

which is the result given by Eq. (2.1a).

In the derivation of Eq. (2.3), we have assumed that $t\leqslant q-t+1$. The result is, however, also correct for $q-t+1\leqslant t$. For example, for $t=q$, the factor (2.3) is $(x_1-\Delta_3+q-h)_h$, which yields correctly the weight space $\mathbb{W}_q^q(\Delta)$ of zeros, which is covered by the lines $x_3=\Delta_3-h+1$ for $h=1,2,\ldots,q$. The modifications of Fig. 2(a) for giving the derivation of the factor (2.1a) when $q-t+1\leqslant t$, but still with $x_3=\Delta_2-h+1$ and $h=1,2,\ldots,t$, is given in Fig. 2(b). For $h\leqslant q-h+1$, the derivation is the same as that given above and gives the factor (2.1a). For $h\geqslant q-h+1$, the factors may be obtained from Fig. 2(b): For $x_1=\Delta_3-q+1,\Delta_3-q+2,\ldots,\Delta_3-t+1$, we have $q-t+1\geqslant 1+d_t(x)=1,2,\ldots,q-t+1$, respectively; for $x_1=\Delta_3-t+1,\Delta_3-t+2,\ldots,\Delta_3-q+h$, we have $1+d_t(x)\leqslant q-t+1$ and the multiplicity of each zero is $q-t+1$; and for $x_1=\Delta_3-q+h$, $\Delta_3-q+h+1,\ldots,\Delta_3-t+h$, we have $q-t+1\geqslant 1+d_t(x)=q-t+1,\ldots,2,1$, respectively. Thus, for $x_3=\Delta_2-h+1$ with $h\geqslant q-h+1$ and $h\leqslant t$, the polynomial $P_q^t(\Delta;x)$ must contain the factors

$$\begin{aligned}
&(x_1-\Delta_3+q-1)(x_1-\Delta_3+q-2)^2\cdots(x_1-\Delta_3+t)^{q-t}\\
&\quad\times(x_1-\Delta_3+t-1)^{q-t+1}(x_1-\Delta_3+t-2)^{q-t+1}\\
&\quad\times\cdots(x_1-\Delta_3+q-h)^{q-t+1}\\
&\quad\times(x_1-\Delta_3+q-h-1)^{q-t}\\
&\quad\times(x_1-\Delta_3+q-h-2)^{q-t-1}\cdots(x_1-\Delta_3+t-h).
\end{aligned}$$

These terms again combine to give exactly the factors (2.1a).

The preceding results prove the lemma for $h=1,2,\ldots,t$ [the factors (2.1a)]. The proof that $P_q^t(\Delta;x)$ contains the factors (2.1b) for $x_3=\Delta_2-h+1$ with $h=t,t+1,\ldots,q$ is carried out similarly.

The converse of the lemma is obvious since the set of zeros of the factors (2.1a) and (2.1b) exactly cover, by construction, the zeros in the set $\mathbb{W}_q^t(\Delta)$, including correct multiplicities. □

Lemma 2.1 can now be used to prove a second important property of polynomials having the weight space $\mathbb{W}_q^t(\Delta)$ of zeros.

Lemma 2.2: Let $P_q^t(\Delta;x)$ denote a polynomial having determinantal symmetry and at least the weight space $\mathbb{W}_q^t(\Delta)$ of zeros. Then $P_q^t(\Delta;x)$ is of total degree at least $2t(q-t+1)$ in x.

Proof: Consider the value of $P_q^t(\Delta;x)$ for $x_3=\Delta_2-t+1$. We find from Lemma 2.1 that $P_q^t(\Delta;x)|_{x_3=\Delta_2-t+1}$ contains the factor

$$\prod_{s=1}^{t}(x_1-\Delta_3+s-1)_{q-t+1}.$$

Since $P_q^t(\Delta;x)$ has determinantal symmetry, it follows that $P_q^t(\Delta;x)|_{x_3=\Delta_2-t+1}$ contains the factors

$$\begin{aligned}
&\prod_{s=1}^{t}(-\Delta_2+s-1)_{q-t+1}(-\Delta_3+s-1)_{q-t+1}\\
&\quad\times(-\Delta_2-x_1+s-1)_{q-t+1}\\
&\quad\times(x_1-\Delta_3+s-1)_{q-t+1}.
\end{aligned}\tag{2.4}$$

This polynomial is of degree $2t(q-t+1)$ in x_1. Hence the total degree of $P_q^t(\Delta;x)$ in x must be at least $2t(q-t+1)$ since otherwise the degree of the polynomial $P_q^t(\Delta;x)|_{x_3=\Delta_2-t+1}$ will be less than $2t(q-t+1)$ in x_1 and cannot contain the polynomial (2.4) as a factor. □

III. UNIQUENESS THEOREM FOR G_q^t

In this section we first establish a result (Theorem 3.1) that, although somewhat specific in its content, forms the basis for proving a uniqueness theorem (Theorem 3.2) for G_q^t. This theorem is significant in that it shows that the properties (i) total degree $2t(q-t+1)$ in x, (ii) determinantal symmetry, and (iii) exactly the weight space $\mathbb{W}_q^t(\Delta)$ of zeros are, in fact, essentially a unique specification of the polynomial.

We begin with the proof of the following important preliminary result.

Theorem 3.1: Let $Q_q^t(\Delta;x)$ denote a polynomial of total degree in x not exceeding $2t(q-t+1)$ and having determinantal symmetry and at least the weight space $\mathbb{W}_q^t(\Delta)$ of zeros. Also, let this polynomial satisfy

$$Q_q^t(\Delta;x)=0,\quad\text{for }\Delta_1=t-1,t,\ldots,q-1,\tag{3.1a}$$

for all Δ_2, $\Delta_3\in\mathbb{Z}_+$ and $x\in\mathbb{M}$. Then

$$Q_q^t(\Delta;x)\equiv 0,\tag{3.1b}$$

for all $\Delta_i\in\mathbb{Z}_+$ and $x\in\mathbb{M}$.

Proof: By assumption, the polynomial $Q_q^t(\Delta;x)$ possesses determinantal symmetry. In particular, it obeys the relation

$$Q_q^t(\Delta_1,\Delta_2,\Delta_3;x)=Q_q^t(\Delta_2-x_3,\Delta_1+x_3,\Delta_3;x),\tag{3.2}$$

for all $x\in L$. Evaluating relation (3.2) at $x_3=\Delta_2-h+1$, we obtain

$$\begin{aligned}
&Q_q^t(\Delta;x)|_{x_3=\Delta_2-h+1}\\
&\quad=Q_q^t(h-1,\Delta_1+\Delta_2-h+1,\Delta_3;x)|_{x_3=\Delta_2-h+1}=0,
\end{aligned}$$

for each $h=t,t+1,\ldots,q$, where we have used Eq. (3.1a) in equating this relation to zero. Thus the polynomial $Q_q^t(\Delta;x)$, if not identically zero, contains the factor $(\Delta_2-x_3-q+1)_{q-t+1}$. Again invoking determinantal symmetry, we conclude that $Q_q^t(\Delta;x)$ has the form given by

$$Q^t_q(\Delta;x) = \prod_{ijk} (\Delta_i - q + 1)_{q-t+1}$$
$$\times (\Delta_i - x_j - q + 1)_{q-t+1}$$
$$\times (\Delta_i + x_k - q + 1)_{q-t+1} Q_t(\Delta;x). \qquad (3.3)$$

For $t = 1$ or 2, necessarily $Q_t(\Delta;x) = 0$ since otherwise the degree $6(q - t + 1)$ in x of the Π_{ijk} factor exceeds the largest degree $2t(q - t + 1)$ of $Q^t_q(\Delta;x)$; hence the theorem is true for $t = 1,2$. For $t \geqslant 3$, $Q_t(\Delta;x)$ is a polynomial in x of total degree not exceeding $2t(q - t + 1) - 6(q - t + 1) = 2(t - 3)(q - t + 1)$. By assumption, $Q^t_q(\Delta;x)$ has at least the weight space $\mathbb{W}^t_q(\Delta)$ of zeros. Removing the zeros of the factors

$$(\Delta_3 - x_1 - q + 1)_{q-t+1}(\Delta_2 - x_3 - q + 1)_{q-t+1}$$

from $\mathbb{W}^t_q(\Delta)$ leaves the weight space $\mathbb{W}^{t-2}_q(\Delta_1 - 1, \Delta_2 - 1, \Delta_3 - 1)$; hence $Q_t(\Delta;t)$ must possess at least this weight space $\mathbb{W}^{t-2}_q(\Delta_1 - 1, \Delta_2 - 1, \Delta_3 - 1)$ of zeros. Therefore, the polynomial $Q_t(\Delta;x)$ must, by Lemma 2.2, be of total degree in x of at least $2(t - 2)(q - t + 3)$, which is *greater* than the maximal degree $2(t - 3)(q - t + 1)$ found above. This contradiction in the degree of $Q_t(\Delta;x)$ can be avoided if and only if $Q_t(\Delta;x) \equiv 0$, which proves the theorem.

We next use Theorem 3.1 to prove a uniqueness theorem.

Theorem 3.2: Up to a multiplicative factor $\alpha^t_q(S)$, where $S = \Delta_1 + \Delta_2 + \Delta_3$, the polynomial $G^t_q(\Delta;x)$ is the unique polynomial that possesses the following properties: (i) total degree $2t(q - t + 1)$ in x, (ii) determinantal symmetry, and (iii) the weight space $\mathbb{W}^t_q(\Delta)$ of zeros.

Proof: The proof of Theorem 3.2 is lengthy. It is by induction, with the following assumption as the starting point. Induction hypothesis: for each $t = 1,2,...,k$ and $k = 1,2,...,q - 1$, the polynomial $G^t_k(\Delta;x)$ is the unique polynomial, up to a multiplicative factor $\alpha^t_k(S)$, that has the weight space $\mathbb{W}^t_k(\Delta)$ of zeros.

We wish to extend this hypothesis to level $k = q$ with $t = 1,2,...,q$. As a first step, we establish the following result, assuming the induction hypothesis: Every polynomial $P^t_q(\Delta;x)$ of total degree $2t(q - t + 1)$ in x that has determinantal symmetry and the weight space $\mathbb{W}^t_q(\Delta;x)$ of zeros is given in terms of $G^t_q(\Delta;x)$ for the particular values

$$\Delta_1 = t - 1, t, ..., q - 1 \qquad (3.4a)$$

by

$$P^t_q(\Delta;x) = \alpha^t_q(S) G^t_q(\Delta;x). \qquad (3.4b)$$

To prove this property, we first apply to P^t_q the symmetry given by relation (3.2) and evaluate at $x_3 = \Delta_2 - h + 1$ for h an arbitrary positive integer. This yields

$$P^t_q(\Delta_1,\Delta_2,\Delta_3;x)\,|_{x_3 = \Delta_2 - h + 1}$$
$$= P^t_q(h - 1, \Delta_1 + \Delta_2 - h + 1, \Delta_3;x)\,|_{x_3 = \Delta_2 - h + 1}. \qquad (3.5)$$

Consider next the weight space $\mathbb{W}^t_q(h - 1, \Delta_1 + \Delta_2 - h + 1, \Delta_3)$ which is that associated with the Δ parameters occurring on the rhs of Eq. (3.5), as shown in Fig. 3. The line $x_3 = \Delta_2 - h + 1$ is seen to lie between the lines

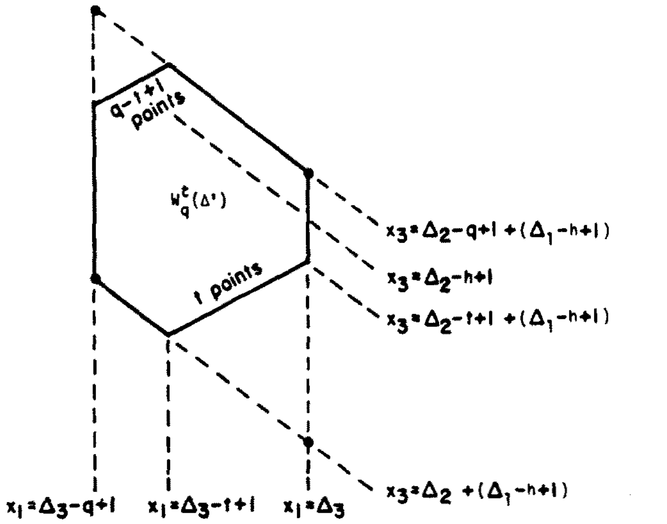


FIG. 3. The weight space $\mathbb{W}^t_q(\Delta')$ for $\Delta' = (h - 1, \Delta_1 + \Delta_2 - h + 1, \Delta_3)$ for each $h \in \mathbb{Z}_+$. The point $(x_1, x_2, \Delta_2 - h + 1) \in \mathbb{W}^t_q(\Delta')$ for all positive integers h and all x_1 such that $\Delta_3 - q + 1 \leqslant x_1 \leqslant \Delta_3$. This property is used in the proof of Theorem 3.2.

$x_3 = (\Delta_1 - h + 1) + \Delta_2 - q + 1$ and $x_3 = (\Delta_1 - h + 1) + \Delta_2 - t + 1$ for all values of Δ_1 given by

$$\Delta_1 = t - 1, t, ..., q - 1$$

(with coincidence of lines for the end values $\Delta_1 = t - 1$ or $q - 1$). Accordingly, the line $x_3 = \Delta_2 - h + 1$ passes through the weight space $\mathbb{W}^t_q(h - 1, \Delta_1 + \Delta_2 - h + 1, \Delta_3)$ for all positive integral values of h. Since by assumption the polynomial $P^t_q(h - 1, \Delta_1 + \Delta_2 - h + 1, \Delta_3;x)$ possesses the weight space $\mathbb{W}^t_q(h - 1, \Delta_1 + \Delta_2 - h + 1, \Delta_3)$ of zeros, it must contain the factor

$$\prod_{s=1}^{t} (x_1 - \Delta_3 + s - 1)_{q - \Delta_1}, \qquad (3.6)$$

for $x_3 = \Delta_2 - h + 1$ (Lemma 2.1 with appropriate changes in notation). Since h is an arbitrary positive integer, we find from Eq. (3.5) that the polynomial $P^t_q(\Delta;x)$ contains the factor (3.6) for each $\Delta_1 = t - 1, t, ..., q - 1$. Since $P^t_q(\Delta;x)$ has determinantal symmetry, it accordingly contains all factors obtained from (3.6) by applying this symmetry. We have thus shown that $P^t_q(\Delta;x)$ has the form given by

$$P^t_q(\Delta;x)$$
$$= \prod_{s=1}^{t} (-\Delta_2 + s - 1)_{q-\Delta_1}(-\Delta_3 + s - 1)_{q-\Delta_1}$$
$$\times (x_1 - \Delta_3 + s - 1)_{q-\Delta_1}$$
$$\times (-x_1 - \Delta_2 + s - 1)_{q-\Delta_1} Q_t(\Delta;x), \qquad (3.7)$$

where $Q_t(\Delta;x)$ is a polynomial of total degree $2t(\Delta_1 - t + 1)$ in x. Relation (3.7) is valid for $\Delta_1 = t - 1, t, ..., q - 1$.

So far we have not used the uniqueness property of the functions $G^t_k(k = 1,2,...,q - 1)$ assumed in the induction hypothesis. This assumption will now be invoked in order to identify the polynomial $Q_t(\Delta;x)$ in relation (3.7). Here the reduction formula given by Eq. (1.10) has the crucial role. We wish to prove that

$$Q_t(\Delta;x) = \alpha(S) G^t_{\Delta_1}(q,\Delta_1+\Delta_2-q,\Delta_1+\Delta_3-q;\; x_1,x_2-\Delta_1+q,x_3+\Delta_1-q), \tag{3.8}$$

for each $t = 0,1,...,\Delta_1$.

Let us recall that the geometrical content of the reduction relation (1.10) is that the removal from $\mathbf{W}^t_q(\Delta)$ of the zeros originating from the linear factors in x_1 in the product

$$\prod_{s=1}^{t} (x_1 - \Delta_3 + s - 1)_{q-\Delta_1} \tag{3.9}$$

is to leave behind exactly the weight space $\mathbf{W}^t_{\Delta_1}(q,\Delta_1+\Delta_2-q,\Delta_1+\Delta_3-q)$ of zeros of

$$G^t_{\Delta_1}(q,\Delta_1+\Delta_2-q,\Delta_1+\Delta_3-q;\; x_1,x_2-\Delta_1+q,x_3+\Delta_1-q) \tag{3.10a}$$

for each point

$$(x_1,x_2-\Delta_1+q,x_3+\Delta_1-q) \in \mathbf{W}^t_{\Delta_1}(q,\Delta_1+\Delta_2-q,\Delta_1+\Delta_3-q). \tag{3.10b}$$

Observing that the factors (3.9) occur also in relation (3.7), we see that removing the zeros of these factors from the weight space $W^t_q(\Delta)$ of zeros of $P^t_q(\Delta;x)$ leaves behind (as above) exactly the weight space $\mathbf{W}^t_{\Delta_1}(q,\Delta_1+\Delta_2-q,\Delta_1+\Delta_3-q)$ of zeros. These zeros must be inherited by the polynomial $Q_t(\Delta;x)$ for each point

$$(x_1,x_2,x_3) \in W^t_q(\Delta;x) - \mathbf{K}^t_q, \tag{3.11}$$

where $\mathbf{K}^t_q$ denotes the set of zeros (with multiplicity) defined by the factors (3.9). From the general definition of the weight space $\mathbf{W}^t_q(\Delta)$ of zeros (see Fig. 1), we verify that the points (x_1,x_2,x_3) given by Eqs. (3.10b) and (3.11) are exactly the same and, moreover, that the multiplicity of each zero is the same.

Since $\Delta_1 < q$ in the polynomial $G^t_{\Delta_1}(\cdots)$ given by (3.10a), this polynomial is, by the induction hypothesis, the unique polynomial [up to a factor $\alpha(S)$] with the weight space (3.10b) of zeros. This result proves relation (3.8).

Substituting $Q_t(\Delta;x)$ from Eq. (3.8) into Eq. (3.7) and comparing the result with the reduction formula (1.10), we obtain Eq. (3.4b), which was to be proved. Using this result and Theorem 3.1, we can now complete the proof of the theorem.

We define the polynomial $Q^t_q(\Delta;x)$ by

$$Q^t_q(\Delta;x) = P^t_q(\Delta;x) - \alpha^t_q(S) G^t_q(\Delta;x), \tag{3.12}$$

where $P^t_q(\Delta;x)$ is a polynomial of degree $2t(q-t+1)$ having determinantal symmetry and the weight space $\mathbf{W}^t_q(\Delta)$ of zeros; thus $P^t_q(\Delta;x)$ satisfies Eqs. (3.4). Then, either $Q^t_q(\Delta;x)$ is identically zero at the outset, or it satisfies the assumptions given in Theorem 3.1; hence, it is identically zero in consequence of the conclusion of that theorem. That $Q^t_q(\Delta;x)$, if not identically zero at the outset, satisfies the assumptions given in Theorem 3.1 may be shown as follows: From the definition (3.12), we find that the degree of $Q^t_q(\Delta;x)$ in x does not exceed $2t(q-t+1)$. Moreover, this polynomial has at least the weight space $\mathbf{W}^t_q(\Delta)$ of zeros, since taking the difference of two polynomials, each of which possesses the weight space $\mathbf{W}^t_q(\Delta)$ of zeros, can at most increase the multiplicity of a zero. Lemma 2.2 implies that the total degree of $G^t_q(\Delta;x)$ in x, when this polynomial is not identically zero, is at least $2t(q-t+1)$. Thus $Q^t_q(\Delta;x)$ is of exactly degree $2t(q-t+1)$, if not identically zero. Finally, Eqs. (3.4) imply that relation (3.1b) of Theorem 3.1 is satisfied. Hence, all the assumptions of Theorem 3.1 are satisfied by the polynomial defined by Eq. (3.12).

We have now proved that

$$P^t_q(\Delta;x) = \alpha^t_q(S) G^t_q(\Delta;x), \tag{3.13}$$

for each $t = 1,2,...,q$; this result extends the induction hypothesis to level q, thus closing the induction loop. This proves Theorem 3.1 provided that the initial step of the induction hypothesis is true. This initial step consists of proving that $G^1_q(\Delta;x)$ is the unique polynomial of total degree $2q$ in x up to a multiplicative factor $\alpha^1_q(S)$, which has the weight space $\mathbf{W}^t_q(\Delta)$ of zeros. This proof was given in Ref. 3, where the polynomial $G_q(\Delta;x) = G^1_q(\Delta;x)$ was first investigated and studied in detail. □

IV. THE POLYNOMIAL $\mathscr{G}^t_q$ AND ITS PROPERTIES

An explicit polynomial form for $G^t_q(\Delta;x)$ was first given as a conjecture in Ref. 6. This polynomial is denoted by $\mathscr{G}^t_q(\Delta;x)$ in Eq. (4.2), where the script letter is a reminder that this is (at the moment) a conjectured form of $G^t_q(\Delta;x)$. A principal result for making this conjecture was the proof in Ref. 3 that

$$G^1_q(\Delta;x) = \mathscr{G}^1_q(\Delta;x). \tag{4.1}$$

The explicit construction of $G^q_q(qqq;x)$ in Ref. 5 and special values of q were also helpful in suggesting the occurrence of the multiplicity numbers $h(\lambda\mu\nu\rho)$ in the definition of $\mathscr{G}^t_q(\Delta;x)$ (see below). Further details motivating this definition are given in Ref. 6. The definition of $\mathscr{G}^t_q(\Delta;x)$ is the following:

$$\mathscr{G}^t_q(\Delta;x) = \mathscr{G}^t_q(A) \equiv (-1)^{t(q-t+1)} \prod_{s=1}^{t} \frac{(q-s+1)!}{(s-1)!} \sum_{\lambda\mu\nu\rho} h(\lambda\mu\nu\rho) \frac{\Pi^t_{s=1}(\Delta_1+\Delta_2+\Delta_3-t-q-s+3)_{\rho_s}}{M(\rho)}$$
$$\times F_{q-t+1,\lambda}(a_{11},a_{12},a_{13}) F_{q-t+1,\mu}(a_{21},a_{22},a_{23}) F_{q-t+1,\nu}(a_{31},a_{32},a_{33}), \tag{4.2}$$

where the a_{ij} denote the entries in the array A given by Eq. (1.6). The other symbols in this result have the following definitions.

(i) The symbol $\lambda = [\lambda_1,\lambda_2,...,\lambda_t]$ denotes an irrep label of $U(t)$, with $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_t \geqslant 0$, each λ_i = a non-negative integer; λ may also be regarded as the *shape* of a Young frame $Y(\lambda)$. The symbols μ, ν,... denote irrep labels of the same type as λ.

(ii) The symbol $h(\lambda\mu\nu\rho)$ denotes the number of times irrep $[q-t+1,...,q-t+1]$ ($q-t+1$ repeated t times) is contained in the direct product $\lambda \times \mu \times \nu \times \rho$ and is defined to be zero if $[q-t+1,...,q-t+1] \notin \lambda \times \mu \times \nu \times \rho$.

(iii) The symbol $M(\lambda)$ denotes the measure of the Young frame $Y(\lambda)$ and has the definition

$$M(\lambda)=\frac{\Pi_{s=1}^{t}(\lambda_s+t-s)!}{\Pi_{r<s}(\lambda_r-\lambda_s+s-r)}=(\mathrm{Dim}\,\lambda)^{-1}\prod_{s=1}^{t}(t-s+1)_{\lambda_s},\quad \mathrm{Dim}\,\lambda=\frac{\Pi_{r<s}(\lambda_r-\lambda_s+s-r)}{1!2!\cdots(t-1)!},\tag{4.3}$$

with corresponding definitions for $M(\mu)$, $M(\nu)$, and $M(\rho)$.

(iv) For each non-negative integer k and each Young frame $Y(\lambda)$, the function $F_{k,\lambda}(x,y,z)$ is defined for indeterminates x,y,z by

$$F_{k,\lambda}(x,y,z)=\begin{cases}[1/M(\lambda)]\Pi_{s=1}^{t}[x+t-s]_{k-\lambda_s}[y+s-1]_{\lambda_s}[z+s-1]_{\lambda_s},\\ 0,\quad \text{if any}\quad \lambda_s>k.\end{cases}\tag{4.4}$$

As remarked in the Introduction, our goal is to prove that $G_q^t(\Delta;x)=\mathscr{G}_q^t(\Delta;x)$, but so far we have been unable to find a direct proof. In place of such a direct proof, we have adopted the strategy of proving that these two functions share so many properties that via a uniqueness theorem, they are necessarily identical. This is a sizeable task, but ultimately successful.

The following features of this approach stand in sharp contrast: While the function $G_q^t(\Delta;x)$ as defined by Eqs. (1.4) is unwieldy and almost intractable, the proof of properties (i)–(iii) in Sec. I—although admittedly tedious and lengthy—turns out to be reasonably straightforward.[1] On the other hand, while the polynomial $\mathscr{G}_q^t(\Delta;x)$ as defined by Eqs. (4.2)–(4.4) is quite comprehensible, the proof of properties (i)–(iii) is deceptively difficult. Indeed, the major hurdle is to find a simple proof of the identity (transpositional symmetry):

$$\mathscr{G}_q^t(A)=\mathscr{G}_q^t(\tilde{A}),\tag{4.5}$$

where $\tilde{A}$ is the array obtained from A by matrix transposition.

In the remainder of this section, we develop some of the properties of $\mathscr{G}_q^t(\Delta;x)$.

We begin with four properties, the first two of which are easy consequences of the definitions (4.2)–(4.4).

(a) The polynomial $\mathscr{G}_q^t$ has total degree $2t(q-t+1)$ in x and total degree $2t(q-t+1)$ in Δ.

(b) The polynomial $\mathscr{G}_q^t$ is invariant under all transformations of $(\Delta_1,\Delta_2,\Delta_3,x_1,x_2,x_3)$ corresponding to row interchanges and column 2–column 3 interchange in the array A.

(c) Under the assumption of transpositional symmetry (4.5), the polynomial $\mathscr{G}_q^t(\Delta;x)$ satisfies exactly the same reduction formulas (1.12) and (1.13) as does the polynomial $G_q^t(\Delta;x)$.

(d) Under the assumption of transpositional symmetry (4.5), the polynomial

$$\mathscr{G}_q^t(A)\,|_{a_{ij}=h-t},\quad h\in\{1,2,\dots,t\},\tag{4.6a}$$

contains the factors

$$\prod_{s=1}^{h}\left(\prod_{\substack{k=1\\k\neq j}}^{3}(-a_{ik}-s+1)_{q-t+1}\times\prod_{\substack{k=1\\k\neq i}}^{3}(-a_{kj}-s+1)_{q-t+1}\right).\tag{4.6b}$$

We need to prove only (c) and (d) since (a) and (b) are obvious.

Proof: The proof of (c) is given directly from the definition (4.2) of $\mathscr{G}_q^t(\Delta;x)$. Since for each positive integer n, the property $[n]_a=0$ holds unless $a\leqslant n$, it follows that

$$F_{q-t+1,\lambda}(a_{11},a_{12},a_{13})=0\tag{4.7}$$

unless $q-t+1-\lambda_t\leqslant a_{11}=\Delta_1-t+1$, that is, unless $\lambda_t\geqslant q-\Delta_1$. Since $\lambda=[\lambda_1,\lambda_2,\dots,\lambda_t]$ is a partition, we thus find that the only terms contributing to the sum in Eq. (4.2) have

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_t\geqslant q-\Delta_1.\tag{4.8}$$

For all such partitions λ, a nontrivial but straightforward calculation shows that

$$\begin{aligned}&F_{q-t+1,\lambda}(a_{11},a_{12},a_{13})\\&\quad=\prod_{s=1}^{t}\frac{(\Delta_1-s+1)!}{(q-s+1)!}\\&\quad\times\left(\prod_{s=1}^{t}[a_{12}+s-1]_{q-\Delta_1}[a_{13}+s-1]_{q-\Delta_1}\right)\\&\quad\times F_{\Delta_1-t+1,\lambda'}(a'_{11},a'_{12},a'_{13}),\end{aligned}\tag{4.9a}$$

where $\lambda'=[\lambda'_1,\lambda'_2,\dots,\lambda'_t]$ is the partition defined in terms of λ by $\lambda'_s=\lambda_s-q+\Delta_1$, $s=1,2,\dots,t$ and a'_{11},a'_{12},a'_{13} are defined by

$$\begin{aligned}&a'_{11}=q-t+1,\quad a'_{12}=a_{12}+\Delta_1-q,\\&a'_{13}=a_{13}+\Delta_1-q.\end{aligned}\tag{4.9b}$$

Since

$$\lambda=\lambda'+[q-\Delta_1,\dots,q-\Delta_1],$$

we find that the summation in definition (4.2) is also reduced to one over all partitions λ', μ, ν, ρ such that $\lambda'\times\mu\times\nu\times\rho\in[\Delta_1-t+1,\Delta_1-t+1,\dots,\Delta_1-t+1]$, which in turn implies that

$$\mu_s\leqslant\Delta_1-t+1,\quad \nu_s\leqslant\Delta_1-t+1,$$

for $s=1,2,\dots,t$. For partitions μ and ν satisfying these conditions, we find

$$\begin{aligned}&F_{q-t+1,\mu}(a_{21},a_{22},a_{23})\\&\quad=[a_{21}+t-s]_{q-\Delta_1}F_{\Delta_1-t+1,\mu}(a'_{21},a_{22},a_{23}),\end{aligned}\tag{4.10a}$$

$$\begin{aligned}&F_{q-t+1,\nu}(a_{31},a_{32},a_{33})\\&\quad=[a_{31}+t-s]_{q-\Delta_1}F_{\Delta_1-t+1,\nu}(a'_{31},a_{32},a_{33}),\end{aligned}\tag{4.10b}$$

where

$$a'_{21}=a_{21}+\Delta_1-q,\quad a'_{31}=a_{31}+\Delta_1-q.\tag{4.10c}$$

Using relations (4.9) and (4.10) in Eq. (4.2), we find upon setting $\Delta_1=h-1$ that

$$\begin{aligned}\mathscr{G}_q^t(A_{11}(h)) &= \mathscr{G}_q^t(A)\,|_{a_{11}=h-t}\\ &= (-1)^{t(q-h+1)}\prod_{s=1}^{t}((-a_{12}-s+1)_{q-h+1}\\ &\quad\times(-a_{13}-s+1)_{q-h+1}(-a_{21}-s+1)_{q-h+1}\\ &\quad\times(-a_{31}-s+1)_{q-h+1})\mathscr{G}_{h-1}^t(A'_{11}(h)),\end{aligned} \tag{4.11a}$$

where

$$A'_{11}(h) = A_{11}(h) + S_{11}(h), \tag{4.11b}$$

for each $h = t, t+1,\ldots,q+1$. In obtaining Eq. (4.11a), we have also used

$$\begin{aligned}&\Delta'_1+\Delta'_2+\Delta'_3-t-(h-1)-s+3\\ &\quad=(h-1)+\Delta_2+\Delta_3-t-q-s+3,\end{aligned} \tag{4.12a}$$

where the Δ'_i are obtained from column 1 of $A'_{11}(h)$ and are given by

$$\begin{aligned}&\Delta'_1 = q,\\ &\Delta'_2=\Delta_2-(q-h+1),\ \Delta'_3=\Delta_3-(q-h+1).\end{aligned} \tag{4.12b}$$

Relation (4.11a) is of the same form as relation (1.10) [see, also, Eqs. (1.11)]. We now apply the row symmetry of $\mathscr{G}_q^t$ and thus extend the validity of Eq. (4.11a) to the form given by Eq. (1.12a) for all indices $(i,j) = (1,1)$, $(2,1)$, $(3,1)$. Under the assumption that $\mathscr{G}_q^t$ also has transpositional symmetry, we obtain the desired proof of (c) since this symmetry together with row symmetry generates determinantal symmetry. □

Proof: The first part of the proof of (d) is similar to that given for property (c). We find that

$$F_{q-t+1,\lambda}(h-t,a_{12},a_{13})=0, \tag{4.13a}$$

for each $h\in\{1,2,\ldots,t\}$ unless the first h parts of the partition λ satisfy

$$\lambda_1=\lambda_2=\cdots=\lambda_h=q-t+1. \tag{4.13b}$$

For such partitions, we have

$$\frac{M(\lambda')}{M(\lambda)}=\prod_{s=1}^{h}\frac{(s-1)!}{(q-s+1)!}\prod_{s=h+1}^{t}[q-t+s-\lambda_s]_h, \tag{4.14a}$$

where

$$\lambda'=(\lambda_{h+1},\lambda_{h+2},\ldots,\lambda_t). \tag{4.14b}$$

This result gives

$$\begin{aligned}&F_{q-t+1,\lambda}(h-t,a_{12},a_{13})\\ &\quad=\prod_{s=1}^{h}\frac{(t-h+s-1)!}{(q-s+1)!}\prod_{s=1}^{h}(-a_{12}-s+1)_{q-t+1}\\ &\quad\quad\times(-a_{13}-s+1)_{q-t+1}\\ &\quad\quad\times F_{q-t+1,\lambda'}(-t,a_{12}+h,a_{13}+h),\end{aligned} \tag{4.15a}$$

for

$$\lambda=[q-t+1,\ldots,q-t+1,\lambda']. \tag{4.15b}$$

Relation (4.15a) shows that $\mathscr{G}_q^t$ evaluated at $a_{11}=h-t$ contains the factors

$$\prod_{s=1}^{h}(-a_{12}-s+1)_{q-t+1}(-a_{13}-s+1)_{q-t+1}, \tag{4.16a}$$

for each $h\in\{1,2,\ldots,t\}$. Unlike the proof of (c), we cannot easily establish directly from the expression for $\mathscr{G}_q^t(A)\,|_{a_{11}=h-t}$ the occurrence of the factors

$$\prod_{s=1}^{h}(-a_{21}-s+1)_{q-t+1}(-a_{31}-s+1)_{q-t+1}. \tag{4.16b}$$

We can, however, apply the assumed transpositional symmetry of $\mathscr{G}_q^t$ to infer that the factors (4.16b) also occur.

The above results for $a_{11}=h-t$ and the determinantal symmetry of $\mathscr{G}_q^t$ now imply property (d). □

The following principal property of the polynomial $\mathscr{G}_q^t(\Delta;x)$ is now easily proved from the results given above.

Theorem 4.1: The properties given by (c) and (d) imply that $\mathscr{G}_q^t(\Delta;x)$ has at least the weight space $\mathbf{W}_q^t(\Delta)$ of zeros.

Proof: The $(i,j)=(3,3)$ case of Eqs. (4.6) and (1.12a), applied to $\mathscr{G}_q^t$ [property (c)], corresponds to $x_3=\Delta_2-h+1$. Equation (4.6b) contains the factors

$$\prod_{s=1}^{h}(-a_{13}-s+1)_{q-t+1} = \prod_{s=1}^{h}(x_1-\Delta_3+t-s)_{q-t+1},$$

for each $h\in\{1,2,\ldots,t\}$, while Eq. (1.12a) contains, on the rhs, the factors

$$\prod_{s=1}^{t}(-a_{13}-s+1)_{q-h+1} = \prod_{s=1}^{t}(x_1-\Delta_3+t-s)_{q-h+1},$$

for each $h\in\{t,t+1,\ldots,q\}$. Thus the polynomial $\mathscr{G}_q^t(\Delta;x)$, evaluated at $x_3=\Delta_2-h+1$, satisfies the conditions of Lemma 2.1 (converse part). □

We can now prove the principal result of this paper.

Theorem 4.2: Under the assumption of transpositional symmetry of $\mathscr{G}_q^t$, the polynomials G_q^t and $\mathscr{G}_q^t$ are identical:

$$G_q^t(\Delta;x)=\mathscr{G}_q^t(\Delta;x). \tag{4.17}$$

Proof: The polynomial $\mathscr{G}_q^t$ is of total degree $2t(q-t+1)$ in x and, under the assumption that $\mathscr{G}_q^t(A)=\mathscr{G}_q^t(\tilde{A})$, has been shown to have determinantal symmetry and at least the weight space $\mathbf{W}_q^t(\Delta)$ of zeros. Thus all the conditions of Theorem 3.2 are satisfied and this implies that

$$\mathscr{G}_q^t(\Delta;x)=\alpha_q^t(S)G_q^t(\Delta;x),$$

where $\alpha_q^t(S)$ is independent of x. Evaluating this relation at, say $x_3=\Delta_2-t+1$, which corresponds to $a_{33}=0$ in Eq. (1.13), and using the fact [property (c)] that

$$\mathscr{G}_q^t(A)\,|_{a_{33}=0}=G_q^t(A)\,|_{a_{33}=0},$$

we find $\alpha_q^t(S)=1$. □

Using Theorem 4.2, we can now derive a second reduction formula for G_q^t in addition to that given by Eqs. (1.12).

The *second general reduction formula* is

$$G_q^t(A_{ij}(h))$$
$$= \mathscr{G}_q^t(A_{ij}(h))$$
$$= (-1)^{h(q-t+1)} \prod_{s=1}^{h} \left(\prod_{\substack{k=1\\k\neq j}}^{3} (-a_{ik}-s+1)_{q-t+1} \right.$$
$$\left. \times \prod_{\substack{k=1\\k\neq i}}^{3} (-a_{kj}-s+1)_{q-t+1} \right)$$
$$\times G_{q-h}^{t-h}(A''_{ij}(h)), \qquad (4.18a)$$

for each $h \in \{1,2,...,t\}$, where $A_{ij}(h)$ is defined by Eq. (1.12b) and $A''_{ij}(h)$ is defined by

$$A''_{ij}(h) = A_{ij}(h) + \Delta_{ij}(h), \qquad (4.18b)$$

in which $\Delta_{ij}(h)$ is the 3×3 shift matrix having entry $-h$ in position (i,j), zeros as entries in the (i,j) minor, and h as entry in the remaining four positions. For example,

$$\Delta_{12}(h) = h\begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \qquad (4.18c)$$

Because of determinantal symmetry, a proof of Eqs. (4.18) may be given by taking a special case, say $(i,j) = (3,3)$, so that $a_{33} = h - t$ implies $x_3 = \Delta_2 - h + 1$. The occurrence of the multiplicative factors in relation (4.18a) is implied by Theorem 4.1, so that it is the occurrence of $(-1)^{h(q-t+1)}G_{q-h}^{t-h}(A''_{11}(h))$ that must be proved. This may be proved by showing that the polynomial in question has total degree $(t-h)$ $(q-t+1)$ and the weight space $\mathbb{W}_{q-h}^{t-h}(h-1,\Delta_2;\Delta_3)$ of zeros, the details of which we omit. These two properties, determinantal symmetry and Theorem 4.2 (applied to G_{q-h}^{t-h}), and relation (1.12a) may now be used to complete the proof of Eqs. (4.18). □

While we will not make direct use of the full reduction relation (4.18a), we have noted it not only for completeness, but also because of the interesting relationships it expresses between the various G_q^t polynomials.

The next relationship we derive is a general one for the polynomial $\mathscr{G}_q^t$ defined by Eq. (4.2), *there being no assumption of transpositional symmetry*. This relation is not only a useful alternative form of Eq. (4.2), but will be used explicitly in Sec. V to derive a necessary condition for transpositional symmetry to be valid.

Theorem 4.2: The polynomial $\mathscr{G}_q^t(\Delta;x)$ defined by Eq. (4.2) can be written in the following form:

$$\mathscr{G}_q^t(\Delta;x)$$
$$= (-1)^{t(q-t+1)} \prod_{s=1}^{t} \frac{(q-s+1)!}{(s-1)!}$$
$$\times \sum_{\lambda\mu\nu} g(\lambda\mu\bar{\nu}) F_{q-t+1,\lambda}(a_{11},a_{12},a_{13})$$
$$\times F_{q-t+1,\mu}(a_{21}, -a_{21}-a_{22}+q-2t+1,$$
$$a_{21}-a_{23}+q-2t+1)$$
$$\times F_{q-t+1,\nu}(a_{31},a_{32},a_{33}), \qquad (4.19a)$$

where $\bar{\nu} = [\bar{\nu}_1,\bar{\nu}_2,...,\bar{\nu}_t]$ is the partition defined by

$$\bar{\nu}_s = q-t+1-\nu_{t-s+1}, \quad s = 1,...,t, \qquad (4.19b)$$

and $g(\lambda\mu\bar{\nu})$ denotes the Littlewood–Richardson number, which is the multiplicity of irrep $\bar{\nu}$ in the direct product $\lambda \times \mu$.

Proof: The key result needed for transforming the definition (4.2) of $\mathscr{G}_q^t(\Delta;x)$ to the form (4.12a) is proved in Ref. 8 and is called *the generalized Saalschütz identity* (for $t = 1$, this identity reduces to the well-known one—see, for example, Bailey[9]). The generalized Saalschütz identity is a general polynomial relation for arbitrary variables x, y, z and may be expressed as

$$\sum_{\rho\mu} g(\rho\mu\kappa)M^{-1}(\rho)$$
$$\times \prod_{s=1}^{t} (x+y+z+t-k-s+1)_{\rho_s} F_{k,\mu}(x,y,z)$$
$$= F_{k,\kappa}(x, -x-y-t+k, -x-z-t+k). \quad (4.20)$$

Relation (4.20) is derived from the generalized Saalschütz identity proved in Ref. 8 [see Eq. (5.4)] by making the following identification of notation:

$$F_{k,\lambda}(x,y,z) = \prod_{s=1}^{t} (x+t-k-s+1)_k$$
$$\times \langle {}_2\mathscr{F}_1(-y,-z;x+t-k)|\lambda\rangle. \qquad (4.21)$$

Using the identity

$$h(\lambda\mu\nu\rho) = \sum_{\kappa} g(\rho\mu\kappa)g(\lambda\nu\bar{\kappa}) \qquad (4.22)$$

and relation (4.20), it is now straightforward to transform the rhs of Eq. (4.2) to that in Eq. (4.19a) [choose $k = q-t+1$, $x = a_{21}$, $y = a_{22}$, $z = a_{23}$ in the identity (4.20), hence $x+y+z = \Delta_1+\Delta_2+\Delta_3-3t+3$].

Remark: The termwise symmetry of $\mathscr{G}_q^t$ in the original form (4.2) under all row permutations of A has been "destroyed" by the transformation (4.20). This row symmetry for $\mathscr{G}_q^t$ in the form (4.19a) is now expressed by the equality of (4.19a) to the two similar forms obtained by applying (4.20) to each of rows 1 and 3 of A, that is, by using the alternative choices $(x,y,z) = (a_{11},a_{12},a_{13})$ or $(a_{31},_{32},a_{33})$ in effecting the transformation (4.20) in Eq. (4.2).

V. EXPRESSIONS FOR $\mathscr{G}_q^t$ IN TERMS OF HYPERGEOMETRIC COEFFICIENTS

Relation (4.21) may be used to express $\mathscr{G}_q^t$, as given by Eqs. (4.2) and (4.19a), in terms of the generalized hypergeometric coefficients defined by (see Refs. 8, 10, and 11)

$$\langle {}_2\mathscr{F}_1(a,b;c)|\lambda\rangle$$
$$= M^{-1}(\lambda) \prod_{s=1}^{t} \frac{(a-s+1)_{\lambda_s}(b-s+1)_{\lambda_s}}{(c-s+1)_{\lambda_s}}$$
$$= \mathrm{Dim}\,\lambda \prod_{s=1}^{t} \frac{(a-s+1)_{\lambda_s}(b-s+1)_{\lambda_s}}{(t-s+1)_{\lambda_s}(c-s+1)_{\lambda_s}}, \qquad (5.1)$$

$$\langle {}_1\mathscr{F}_0(a)|\lambda\rangle = M^{-1}(\lambda) \prod_{s=1}^{t} (a-s+1)_{\lambda_s}$$
$$= \mathrm{Dim}\,\lambda \prod_{s=1}^{t} \frac{(a-s+1)_{\lambda_s}}{(t-s+1)_{\lambda_s}}. \qquad (5.2)$$

Here a, b, c may be arbitrary complex parameters ($c\neq t-1, t-2,\ldots$). Since it is the coefficients (5.1) and (5.2) that occur directly in the definitions of the generalized Gauss hypergeometric functions given in Refs. 8 and 11, it is useful to formulate the polynomials $\mathscr{G}_q^t$ directly in terms of these coefficients. This will allow us to formulate the transpositional symmetry of $\mathscr{G}_q^t$ as a property of the hypergeometric coefficients.

Using relation (4.21) in Eq. (4.2), we obtain the following expression for the polynomial $\mathscr{G}_q^t$:

$$\begin{aligned}\mathscr{G}_q^t(A) = \prod_{s=1}^{t} \frac{(q-s+1)!}{(s-1)!} &\times \prod_{i=1}^{3}\prod_{s=1}^{t}(-a_{i1}-s+1)_{q-t+1}\\ &\times \sum_{\lambda\mu\nu\rho} h(\lambda\mu\nu\rho)\langle {}_1\mathscr{F}_0(K-l)|\rho\rangle\\ &\times\langle {}_2\mathscr{F}_1(-a_{12},-a_{13};a_{11}-l)|\lambda\rangle\\ &\times\langle {}_2\mathscr{F}_1(-a_{22},-a_{23};a_{21}-l)|\mu\rangle\\ &\times\langle {}_2\mathscr{F}_1(-a_{32},-a_{33};a_{31}-1)|\nu\rangle,\end{aligned} \tag{5.3a}$$

where we have defined

$$l=q-2t+1,\quad K=\text{magic square parameter of } A. \tag{5.3b}$$

The generalized Saalschütz identity proved in Ref. 8 is

$$\begin{aligned}&\sum_{\rho\sigma} g(\rho\sigma\kappa)\langle {}_1\mathscr{F}_0(c-a-b)|\rho\rangle\langle {}_2\mathscr{F}_1(a,b;c)|\sigma\rangle\\ &\quad=\langle {}_2\mathscr{F}_1(c-a,c-b;c)|\kappa\rangle.\end{aligned} \tag{5.4}$$

Using relation (5.4) for

$$(a,b,c)=(-a_{i2},-a_{i3},a_{i1}-l), \tag{5.5}$$

$i=1$, 2, or 3 with $\sigma=\lambda$, μ, or ν, respectively, and relation (4.22) for the appropriate permutation of (λ,μ,ν) [$h(\lambda\mu\nu\rho)$ is symmetric in the partitions λ, μ, ν, ρ], we can transform Eq. (5.3a) to any one of three possible new forms. We choose the case $i=1$, set $h(\lambda\mu\nu\rho) =\Sigma_\kappa g(\rho\mu\kappa)g(\nu\lambda\bar{\kappa})$, effect the transformation (5.5) with $\sigma=\mu$, rename dummy summation partitions by replacing ν by μ, κ by $\bar{\nu}$, and use $g(\mu\lambda\bar{\nu})=g(\mu\nu\bar{\lambda})$; we thus bring $\mathscr{G}_q^t$ of Eq. (5.3a) to the following form:

$$\begin{aligned}\mathscr{G}_q^t(A) = \prod_{s=1}^{t} \frac{(q-s+1)!}{(s-1)!} &\times \prod_{i=1}^{3}\prod_{s=1}^{t}(-a_{i1}-s+1)_{q-t+1}\\ &\times\sum_{\lambda}\langle {}_2\mathscr{F}_1(-a_{12},-a_{13};a_{11}-l)|\rangle\sum_{\mu\nu} g(\mu\nu\bar{\lambda})\\ &\times\langle {}_2\mathscr{F}_1(-a_{32},-a_{33};a_{31}-l)|\mu\rangle\\ &\times\langle {}_2\mathscr{F}_1(a_{21}+a_{22}-l,a_{21}+a_{23}-l;a_{21}-l)|\nu\rangle.\end{aligned} \tag{5.6}$$

The next step in bringing $\mathscr{G}_q^t$ to a new form is motivated by transpositional symmetry. For this we define the new variables a, b, c, d, e by

$$\begin{aligned}&a=-a_{33},\quad b=-a_{32},\quad d=-a_{22},\\ &e=-a_{23},\quad c=K-l.\end{aligned} \tag{5.7}$$

We also define for each partition $\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_t)$ the function A_λ by

$$\begin{aligned}A_\lambda\begin{pmatrix}a,b,d,e\\ c\end{pmatrix} &= \prod_{s=1}^{t}(a+b+c-s+1)_{\lambda_s}(d+e+c-s+1)_{\lambda_s}\\ &\times\sum_{\mu\nu} g(\mu\nu\lambda)\langle {}_2\mathscr{F}_1(a,b;a+b+c)|\mu\rangle\\ &\times\langle {}_2\mathscr{F}_1(d+c,e+c;d+e+c)|\nu\rangle.\end{aligned} \tag{5.8}$$

Combining definition (5.8) and Eq. (5.6) and carrying out some simplifying algebraic steps, we find the following expression for $\mathscr{G}_q^t$ in terms of the A_λ functions:

$$\begin{aligned}\mathscr{G}_q^t(A) = (-1)^{t(q-t+1)}\left[\prod_{s=1}^{t}\frac{(q-s+1)!}{(s-1)!}\right]\sum_{\lambda} M^{-1}(\lambda)\Bigg[\prod_{s=1}^{t}(-1)^{\lambda_s}(-a_{11}-s+1)_{q-t+1-\lambda_s}(-a_{12}-s+1)_{\lambda_s}\\ \times(-a_{13}-s+1)_{\lambda_s}(-a_{21}-s+1)_{\lambda_s}(-a_{31}-s+1)_{\lambda_s}\Bigg]A_{\bar{\lambda}}\begin{pmatrix}a,b,d,e\\ c\end{pmatrix}.\end{aligned} \tag{5.9a}$$

The array A [see Eq. (1.6)] is expressed in terms of the variables a, b, c, d, e defined in Eqs. (5.7) by

$$A=\begin{pmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & -d & -e\\ a_{31} & -b & -a\end{pmatrix}, \tag{5.9b}$$

with

$$\begin{aligned}a_{11}&=-(c+l)-(a+b+d+e),\\ a_{12}&=(c+l)+(b+d),\\ a_{13}&=(c+l)+(a+e),\\ a_{21}&=(c+l)+(d+e),\\ a_{31}&=(c+l)+(a+b),\end{aligned} \tag{5.9c}$$

where we recall that $l=q-2t+1$.

The summation in Eq. (5.9a) is over all partitions λ such that

$$q-t+1\geqslant\lambda_1\geqslant\cdots\geqslant\lambda_t\geqslant 0. \tag{5.9d}$$

It is the functions $A_{\bar{\lambda}}$, which are defined by Eq. (5.8) for arbitrary partition λ, hence also for the conjugate partitions $\bar{\lambda}$, that occur in Eq. (5.9a).

The following theorem is now obvious from the form of $\mathscr{G}_q^t$ given by Eq. (5.9a).

Theorem 5.1: A sufficient condition for transpositional symmetry of the polynomial $\mathscr{G}_q^t$, that is, for

$$\mathscr{G}_q^t(A)=\mathscr{G}_q^t(\tilde{A}), \tag{5.10a}$$

is that the polynomials

$$A_\lambda \begin{pmatrix} a,b,d,e \\ c \end{pmatrix} \tag{5.10b}$$

be invariant under the interchange of b and e.

Remark: We have not proved that the symmetry of the function A_λ under the interchange of b and e is *necessary* for transpositional symmetry of $\mathscr{G}_q^t$. This result is not immediately evident from Eq. (5.9a). Let us note, however, that for, say $d = 0$, we can use the Saalschütz identity (5.4) (for appropriate parameter identification) in the definition (5.8) to obtain

$$A_\lambda \begin{pmatrix} a,b,0,e \\ c \end{pmatrix} = M^{-1}(\lambda) \prod_{s=1}^{t} (a+c-s+1)_{\lambda_s} \times (b+c-s+1)_{\lambda_s} (e+c-s+1)_{\lambda_s}, \tag{5.11a}$$

which is symmetric under all permutations of a, b, e, hence under the interchange of b and e. It is also trivial to prove

$$A_\lambda \begin{pmatrix} 0,b,d,e \\ c \end{pmatrix} = M^{-1}(\lambda) \prod_{s=1}^{t} (b+c-s+1)_{\lambda_s} \times (d+c-s+1)_{\lambda_s} (e+c-s+1)_{\lambda_s}. \tag{5.11b}$$

These special results, and others, suggest the validity of the general symmetry of A_λ under the interchange of b and e, but this general result is very difficult and lengthy to prove and requires methods quite unlike those of the present paper. Because of this, the proof will be given separately (Ref. 7).

VI. CONCLUDING REMARKS

The construction of the denominator function $D^2(\Gamma_t,x)$ is a major step toward a complete algebraic determination of the matrix elements of all canonical unit tensor operators (WCG coefficients). Not only is the denominator function conceptually important—in the form of the map: tensor operators → invariant norm (denominator function)—which is itself equivalent to the canonical resolution of the multiplicity, but the denominator function also enters both in the WGC coefficients themselves and in the construction of Racah operators. In fact, all Racah coefficients on the boundary (maximal shifts) are simply square roots of fractions using the appropriate denominator functions.

It is useful to note that the construction of the canonical multiplicity splitting in I achieved at the same time an explicit construction of exactly those Racah functions that effect the canonical splitting. It follows (using the pattern calculus for the elementary operators) that an explicit construction of the set of canonical unit tensor operators is now at hand.

[1] L. C. Biedenharn, M. A. Lohe, and J. D. Louck, "On the denominator function for canonical SU(3) tensor operators," J. Math. Phys. **26**, 1458 (1985). Note that the rhs of Eq. (1.36) lacks a U(2) coefficient. Alternatively, taking the limit $m_{23} \to \infty$ on the lhs inserts $\delta_{\Gamma_{11}\Gamma'_{11}}$.

[2] L. C. Biedenharn, J. D. Louck, E. Chacon, and M. Ciftan, "On the structure of the canonical tensor operators in the unitary groups. I. An extension of the pattern calculus rules and the canonical splitting in U(3)," J. Math. Phys. **13**, 1957 (1972).

[3] L. C. Biedenharn and J. D. Louck, "On the structure of the canonical tensor operators in the unitary groups. II. The tensor operators in U(3) characterized by maximal null space," J. Math. Phys. **13**, 1985 (1972).

[4] J. D. Louck, M. A. Lohe, and L. C. Biedenharn, "Structure of the canonical U(3) Racah functions and the U(3):U(2) projective functions," J. Math Phys. **16**, 2408 (1975).

[5] M. A. Lohe, L. C. Biedenharn, and J. D. Louck, "Structural properties of the self-conjugate SU(3) tensor operators," J. Math. Phys. **18**, 1883 (1977).

[6] L. C. Biedenharn, R. A. Gustafson, M. S. Lohe, J. D. Louck, and S. C. Milne, "Special functions and group theory in theoretical physics," in *Special Functions: Group Theoretical Aspects and Applications*, edited by R. A. Askey, T. H. Koornwinder, and W. Schempp (Reidel, Dordrecht, 1984).

[7] L. C. Biedenharn, A. Bincer, M. A. Lohe, and J. D. Louck (to be submitted for publication).

[8] J. D. Louck and L. C. Biedenharn, "A generalization of the Gauss hypergeometric series," J. Math. Anal. Appl. **59**, 423 (1977).

[9] W. N. Bailey, *Generalized Hypergeometric Series* (Cambridge U. P., London, 1935).

[10] D. P. Shukla, "Certain transformations of generalized hypergeometric series," Ind. J. Pure Appl. Math. **12**, 994 (1981).

[11] L. C. Biedenharn and J. D. Louck, "Some properties of generalized hypergeometric coefficients," Adv. Appl. Math. (accepted for publication).

The inverse of a Borel summable function

W. Boenkost, F. Constantinescu, and U. Scharffenberger
Institut für Angewandte Mathematik, Universität Frankfurt am Main, D-6000 Frankfurt, Federal Republic of Germany

(Received 17 July 1987; accepted for publication 30 December 1987)

Let B_1 be the Borel transform of f, $f(0)\neq 0$. By means of an integral equation similar to that proposed by 't Hooft [*The Whys of Subnuclear Physics*, edited by A. Zichichi (Plenum, New York, 1979)] an explicit representation of the Borel transform B_2 of $1/f$ as a series in convolution powers of B_1 and its derivative is given. It is proved that the singularities of B_1, B_2 on the circle of convergence coincide and have the same strength when approaching these singularities on a path inside their common analyticity domain, but possibly outside the circle of convergence of their power series representation (Borel circle).

I. INTRODUCTION

Borel summability has proved to be a powerful tool in singular perturbation problems of quantum mechanics. The method relies on the following well-known result.

Theorem 1 (Watson–Nevanlinna–Sokal): Let $C_R = \{z\in\mathbb{C}\colon \operatorname{Re} 1/z > 1/R\}$ and let $f\colon \overline{C}_R \to \mathbb{C}$ obey the following conditions.

(i) We note that f is analytic in C_R and continuous in $\overline{C}_R$.

(ii) We note that f possesses an asymptotic expansion

$$f(z) = \sum_{n=0}^{N-1} a_n z^n + R_N(z), \tag{1}$$

with the remainder estimate

$$|R_N(z)| \leqslant A\sigma^N N!|z|^N \tag{2}$$

uniformly in C_R.

Then the Borel transform

$$B(t) = \sum_{n=0}^{\infty} \frac{a_n}{n!} t^n \tag{3}$$

converges (at least) in the circle $\{t\in\mathbb{C}\colon |t| < 1/\sigma\}$ and has an analytic continuation to the region $S(\sigma) = \{t\colon \operatorname{dist}(t,\mathbb{R}_+) < 1/\sigma\}$ satisfying the bound (for all $R' < R$)

$$|B(t)| \leqslant \text{const} \exp(|t|/R') \tag{4}$$

uniformly in every $S(\sigma')$, with $\sigma' > \sigma$.

Furthermore, f is given by the absolutely convergent integral

$$f(z) = \frac{1}{z}\int_0^\infty \exp\left(-\frac{t}{z}\right) B(t)\, dt, \tag{5}$$

valid for all $z\in C_R$. For a proof see Refs. 1 and 2.

Note that the singularity structure of B reveals a large amount of information about f. We want to mention applications in quantum field theory and statistical mechanics. For a recent application to disordered systems see Sec. III.

In some problems the function f can be written as $f(z) = f_1(z)/f_2(z)$, where both f_1 and f_2 obey the hypothesis of Nevanlinna's theorem. Among these problems we mention the eigenvalue perturbation for Schrödinger operators (see Ref. 3), lattice field theories, and disordered systems (see Sec. III for a discussion of this point). It is an easy exercise in formal series manipulation [using the inequality $\sum_{n=0}^{N-1} n!\,(N-n)! \leqslant N!$] to show that, provided $f(0)\neq 0$, $1/f$ also obeys Nevalinna's theorem, i.e., $1/f$ is Borel summable if f has this property. Another proof follows from the Appendix in Ref. 4.

We first give a different proof based on the use of an integral equation; the study of this integral equation will be the main contribution of this paper. The first to propose an integral equation for the Borel transform of $1/f$ was 't Hooft,[5] who used the Laplace transform instead of our $1/z$-multiplied version in (5). Consequently, 't Hooft's[5] integral equation contains a δ-function inhomogeneity stemming from the fact that $f(z) = O(z)$ at the origin and $1/f$ has a pole at the origin. We find it convenient to modify the 't Hooft integral equation as follows.

II. MAIN RESULTS

Suppose that

$$f(z) = \frac{1}{z}\int_0^\infty e^{-t/z} B_1(t)\, dt, \tag{6}$$

where B_1 is supposed to be known. We are searching for a function B_2 obeying

$$\frac{1}{f(z)} = \frac{1}{z}\int_0^\infty e^{-t/z} B_2(t)\, dt. \tag{7}$$

Provided that (7) holds,

$$\begin{aligned} 1 = \frac{f(z)}{f(z)} &= z^{-2}\int_0^\infty e^{-t/z}B_1(t)\,dt \int_0^\infty e^{-t/z}B_2(t)\,dt \\ &= z^{-2}\int_0^\infty e^{-t/z}(B_1 * B_2)\,(t)\,dt, \end{aligned} \tag{8}$$

where

$$(B_1 * B_2)\,(t) = \int_0^t B_1(t-t')B_2(t')\,dt'. \tag{9}$$

To establish (8) we used the convolution theorem for Laplace transforms.[6]

Now an integration by parts yields

$$1 = -z^{-1}e^{-t/z}(B_1 * B_2)\,(t)\,\big|_0^\infty + z^{-1}\int_0^\infty e^{-t/z}(B_1 * B_2)'(t)\,dt. \tag{10}$$

with the first term on the right-hand side 0 because $(B_1 * B_2) \times (0) = 0$. Comparing Borel transforms on both sides we obtain

$$1 = \frac{d}{dt}(B_1 * B_2)\ (t)$$

$$= B_1(0)B_2(t) + \int_0^t B_1'(t-t')B_2(t')\,dt', \qquad (11)$$

which gives the desired integral equation for B_2:

$$B_2(t) = \frac{1}{B_1(0)} - \frac{1}{B_1(0)} \int_0^t B_1'(t-t')B_2(t')\,dt'. \qquad (12)$$

Note that $B_1(0) = f(0) \neq 0$, so that (12) makes sense.

Fortunately a formal solution of Eq. (12) can be given immediately. It is easy to check that

$$\begin{aligned} B_2(t) &= \frac{1}{B_1(0)} - \frac{1}{B_1(0)^2}(B_1(t) - B_1(0)) \\ &\quad + \frac{1}{B_1(0)^3}(B_1(t) - B_1(0)) * B_1'(t) \\ &\quad - \frac{1}{B_1(0)^4}(B_1(t) - B_1(0)) * B_1'(t) * B_1'(t) \\ &\quad + \cdots \end{aligned} \qquad (13)$$

solves (12).

The following theorem shows that the series (13) is well defined.

Theorem 2: Let f obey the hypothesis of Theorem 1 and suppose that $f(0) \neq 0$. Then the series (13) converges in $S(\sigma)$ and represents an analytic function there. Moreover, there exists a $R'' > 0$ such that

$$|B_2(t)| \leqslant \text{const} \exp(|t|/R'') \qquad (14)$$

uniformly in every $S(\sigma')$, with $\sigma' > \sigma$. As a consequence the Laplace transform of B_2 exists and represents a function analytic in $C_{R''}$.

Proof: It is well known[7] that B_1' also satisfies the bound (4). Now choose $\sigma' > \sigma$ and $t \in S(\sigma')$. From Nevanlinna's theorem it is clear that each term of the series represents an analytic function in $S(\sigma')$. The convolutions are performed by taking the integrations along the straight line joining 0 and t. Thus we will estimate

$$\left| \int_0^t dt_1 (B_1(t-t_1) - B_1(0)) \int_0^{t_1} dt_2\, B_1'(t_1 - t_2) \times \cdots \times \int_0^{t_{n-1}} dt_n\, B_1'(t_n) \right|. \qquad (15)$$

We make three crucial observations.

(i) We note that B_1 and B_1' obey the exponential bound (4).

(ii) On the contour of integration $|t_i - t_{i+1}| = |t_i| - |t_{i+1}|$, so

$$\prod_{i=0}^{n} \exp(R'^{-1}|t_i - t_{i+1}|) = \exp(R'^{-1}|t|),$$

where we have set $t_0 = t$ and $t_{n+1} = 0$.

(iii) $$\int_0^{|t|} dt_1 \int_0^{|t_1|} dt_2 \cdots \int_0^{|t_{n-1}|} dt_n = \frac{|t|^n}{n!}.$$

From (i)–(iii) it follows easily that (15) is bounded by

$$\text{const}^n \exp(|t|/R')|t|^n/n!. \qquad (16)$$

Inserting (16) into (13) shows that the series converges uniformly in every compact subset of $S(\sigma)$ and gives an exponential bound (14) uniformly in every $S(\sigma')$ ($\sigma' > \sigma$). The analyticity follows from Vitali's theorem. ∎

We now have the following theorem.

Theorem 3: The function $B_2(t)$ given by (12) or (13) is the Borel transform of $1/f$, i.e.,

$$\frac{1}{z}\int_0^\infty \exp\left(-\frac{t}{z}\right) B_2(t)\,dt = \frac{1}{f(z)}. \qquad (17)$$

Proof: The bound (14), combined with Lebesgue's dominated convergence theorem, shows that for $z \in C_{R''}$ the series (13) can be Laplace transformed term by term to yield the Laplace transform of B_2. Thus we have to study the Laplace transform of the nth term of (13). Using the convolution theorem we obtain

$$\begin{aligned} &\frac{1}{z}\int_0^\infty e^{-t/z}((B_1(\cdot) - B_1(0)) * B_1' * \cdots * B_1')(t)\,dt \\ &= \left(\frac{1}{z}\int_0^\infty e^{-t/z} B_1(t)\,dt - B_1(0)\right) \times \left(\int_0^\infty e^{-t/z} B'(t)\,dt\right)^{n-1} \\ &= (f(z) - B_1(0))^n. \end{aligned} \qquad (18)$$

The evaluation of the second set of parentheses was performed by an integration by parts. Now recall that $B_1(0) = a_0$ and according to the asymptotic expansion (1) $f(z) - B_1(0)$ is $O(z)$ for $z \to 0$ in $C_{R''}$. Thus we choose $|z| \ll 1$ in order to have $|f(z) - B_1(0)| < |B_1(0)|$. Then

$$\frac{1}{z}\int_0^\infty e^{-t/z} B_2(t)\,dt = \frac{1}{B_1(0)} \sum_{n=0}^\infty \left(\frac{f(z) - B_1(0)}{B_1(0)}\right)^n = \frac{1}{f(z)},$$

which was to be proved. ∎

Remarks: (i) The domain of analyticity of B_1 could be larger than that indicated by Theorem 1. Suppose B_1 is analytic inside a circle $C(\Sigma^{-1})$ in the t plane centered at the origin with radius $\Sigma^{-1} > \sigma^{-1}$ and that it cannot be analytically continued to a circle with radius $\Sigma_1^{-1} > \Sigma^{-1}$. Our proof of Theorem 2 actually shows that B_2 is also analytic in $C(\Sigma^{-1})$. A symmetry argument in f and $1/f$ shows that B_2 cannot be analytic in a circle $C(\Sigma_1^{-1})$ with $\Sigma_1^{-1} > \Sigma^{-1}$. The functions B_1 and B_2 have the same set of singularities in their common circle of convergence.

(ii) The series obtained by differentiating (13) term by term also converges uniformly on every compact subset of $S(\sigma)$ and so represents $B_2'(t)$ there. We have

$$\begin{aligned} B_2'(t) &= -[1/B_1(0)^2]B_1'(t) \\ &\quad + [1/B_1(0)^3](B_1' * B_1')(t) \\ &\quad - [1/B_1(0)^4](B_1' * B_1' * B_1')(t) + \cdots. \end{aligned} \qquad (19)$$

The series (19) will help us to study the strength of the singularities of B_2. Since convolution powers of a function are more regular than the function itself we can expect that the first term of series (19) gives us the leading singularity of

B_2'. Under fairly general circumstances this picture is confirmed by the flowing theorem. Here we restrict ourselves to the study of the singularities lying on the circle of convergence of the power series representation (3) of B_2.

Theorem 4: Let ζ be a singularity of B_1' lying on the circle of convergence of (3). Furthermore, suppose that for some $C>0$

$$\int_0^{|t|} |B_1'(t')|\, d|t'| \leqslant C|B_1(t)|\,. \tag{20}$$

Then

$$B_2'(t) = -B_1'(t)/B_1(0)^2 + O(B_1(t)) \tag{21}$$

along the ray joining 0 and ζ.

Proof: For performing the proof we again have to estimate the multiple convolutions constituting series (19). When taking absolute values we treat $t, t_1, \ldots$ as positive real variables to simplify the notation, although they should correctly be denoted by $|t|$, $d|t|$, etc. (for detailed proofs see Ref. 8).

First consider four points: $0 \leqslant t_i \leqslant t_{i-1} \leqslant t_j \leqslant t_{j-1} \leqslant t < \zeta$. (See Fig. 1.) If $t_{i-1} - t_i > t/2$ then necessarily $t_{j-1} - t_j \leqslant t/2$.

Thus we have the estimate

$$\begin{aligned}|B_1'(t_{j-1} - t_j)B_1'(t_{i-1} - t_i)|\\ \leqslant C_1(|B_1'(t_{j-1} - t_j)| + |B_1'(t_{i-1} - t_i)|),\end{aligned} \tag{22}$$

where $C_1 = \sup\{|B_1'(t)|:\ 0 \leqslant t \leqslant \frac{1}{2}(\zeta + \epsilon)\}$ for some $\epsilon > 0$.

With these preliminaries we can estimate the convolution powers of B_1',

$$\begin{aligned}
&\left|\int_0^t dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n\, B_1'(t-t_1)\cdots B_1'(t_{n-1}-t_n)B_1'(t_n)\right|\\
&\leqslant \int_0^t dt_1 \cdots \int_0^{t_{n-1}} dt_n |B_1'(t-t_1)|\cdots|B_1'(t_n)|\\
&\leqslant 2^{n+1}C_1^n \sum_{i=1}^{n+1} \int_0^t dt_1 \cdots \int_0^{t_{i-1}} dt_i |B'(t_{i-1}-t_i)| \int_0^{t_i} dt_{i+1} \cdots \int_0^{t_{n-1}} dt_n\\
&= 2^{n+1}C_1^n \sum_{i=1}^{n+1} \int_0^t dt_1 \cdots \int_0^{t_{i-1}} dt_i |B_1'(t_{i-1}-t_i)| \left|\frac{t_i^{n-i+1}}{(n-i+1)!}\right|\\
&\leqslant 2^{n+1}C_1^n \sum_{i=1}^{n+1} \int_0^t dt_1 \cdots \int_0^{t_{i-2}} dt_{i-1} \frac{t_{i-1}^{n-i+1}}{(n-i+1)!} \int_0^{t_{i-1}} dt_i |B_1'(t_{i-1}-t_i)|\\
&\leqslant 2^{n+1}C_1^n \sum_{i=1}^{n+1} \int_0^t dt_1 \cdots \int_0^{t_{i-2}} dt_{i-1} \frac{t_{i-1}^{n-i+1}}{(n-i+1)!} \int_0^t dt_i |B_1'(t_i)|\\
&= 2^{n+1}C_1^n \sum_{i=1}^{n+1} \frac{t^n}{n!} \int_0^t dt_i |B_1'(t_i)| \leqslant C(n+1)2^{n+1}C_1^n \frac{t^n}{n!}|B_1(t)|\,.
\end{aligned} \tag{23}$$

The bound (23) makes series (19) absolutely convergent, which proves the theorem immediately. The term $-B_1'(t)/B_1(0)^2$ derives from the first term in (19) (without convolution). ■

Remarks: In applications it is often interesting that the asymptotic estimate (21) is not only valid radially, but in an angular sector with vertex at t lying inside the domain of analyticity. The proof of Theorem 4 shows that (21) is true uniformly in a sector with the vertex at t, having an opening angle $2\pi - \alpha$, $\alpha > 0$, if

$$\int_{C_t} |B_1'(t')|\, d|t'| \leqslant C|B_1(t)| \tag{24}$$

is true uniformly for any ray C_t lying in this sector and ending at t. This enables working "outside the Borel circle" (see Refs. 9–11 for an application). There is a large class of functions B_1 satisfying (20). In the following lemma we take $\zeta = -1$.

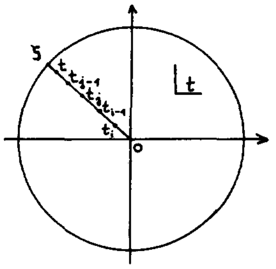


FIG. 1. The four points $0 \leqslant t_i \leqslant t_{i-1} \leqslant t_j \leqslant t_{j-1} \leqslant t < \zeta$ are shown.

Lemma 5: Suppose that (for $\alpha, \beta \geqslant 0$)

$$B_1(t) = [F(t)/(1+t)^\alpha]\log^\beta(1+t)\,, \tag{25}$$

with F analytic in a neighborhood of -1. Then for t near -1 and $|\arg t| < \pi - \epsilon$, $\epsilon > 0$, B_1 obeys

$$\int_{C_t} |B_1'(t')|\, d|t'| \leqslant C|B_1(t)| \tag{26}$$

uniformly for every ray C_t from t_0 to t, with t_0 near t and $|\arg(t - t_0)| \leqslant \pi - \epsilon$.

Proof: For t near -1 we can write

$$|B_1'(t)| \leqslant \text{const}\left(\left|\frac{\log^\beta(1+t)}{(1+t)^\alpha}\right| + \sum_{\gamma=\beta-1}^{\beta} \left|\frac{\log^\gamma(1+t)}{(1+t)^{\alpha+1}}\right|\right). \tag{27}$$

The major contribution in (26) comes from the last term in (27). However (for t near -1),

$$\int_{C_t}\left|\frac{\log^\beta(1+t')}{(1+t')^{\alpha+1}}\right|d|t'|$$

$$\leqslant \sup_{t'\in C_t}|\log^\beta(1+t')|\int_{C_t}\left|\frac{1}{(1+t')^{\alpha+1}}\right|d|t'|$$

$$\leqslant \text{const}|\log^\beta(1+t)|\int_{C_t}\left|\frac{1}{(1+t')^{\alpha+1}}\right|d|t'|$$

$$\leqslant \text{const}|B_1(t)|\,. \qquad \blacksquare$$

A particular case of this lemma appears in Ref. 10.

III. GENERAL REMARKS AND CONCLUSIONS

Following an idea of 't Hooft[5] we study an integral equation for the Borel transform of $1/f$, $f(0)\neq 0$. Because (over the convolution theorem for the Laplace transform) the Borel transform of a product $f\cdot g$ can be computed easily we are able to control the singularity structure of the Borel transform for a quotient such as f/g, $g(0)\neq 0$. By similar methods the log function is studied in Ref. 8.

Besides possible applications to the Borel summability in quantum mechanics, which we do examine here, we have in mind applications in some areas of mathematical physics which make use of the cluster expansion. For the simplest case of a lattice model with continuous spin distribution the cluster expansion (in the high temperature region) typically provides us with an expression of the form $\Sigma_G \Pi_{i\in G}\Phi_i$, where the summation goes over some graphs. Again, Φ_i is typically a quotient of the form f/g, where usually there is good information about the Borel transforms of f and g, including the singularity structure outside the Borel circle. We write the cluster expansion in the "Borel variable." The singularity structure in the Borel plane (mainly *outside* the Borel circle—see, for instance, Refs. 9–11) is responsible for interesting physical properties of the system, such as instantons, etc. The singularity structure in the Borel plane is also the object of interest in a modern branch of mathematics: the theory of resurgent functions with deep application to the theory of differential equations and other areas of interest (see Ref. 12). Applications to the physics of the program described above are given in Refs. 9–11.

At present we are not able to treat more interesting problems such as the identification of the responsibility of instantons for Lifshitz tails in the case of the disordered systems with bounded single-site distribution.

[1] F. Nevanlinna, "Zur Theorie der asymptotischen Potenzreihen," Ann. Acad. Sci. Fenn. Ser. A **12** (3) (1918/19).

[2] A. Sokal, "An improvement of Watson's theorem on Borel summability," J. Math. Phys. **21**, 261 (1980).

[3] B. Simon, "Large orders and summability of eigenvalue perturbation theory: A mathematical overview," Int. J. Quantum Chem. **21**, 3 (1982).

[4] S. Graffi, V. Grecchi, E. Harrell, and H. Silverstone, "The $1/R$ expansion for H_2^+: analyticity, summability and asymptotics," Ann. Phys. **165**, 441 (1985).

[5] G. 't Hooft, in *The Whys of Subnuclear Physics*, edited by A. Zichichi (Plenum, New York, 1979).

[6] D. V. Widder, *The Laplace Transform* (Princeton U. P., Princeton, NJ, 1941).

[7] G. Auberson and G. Mennessier, "Some properties of Borel summable function," J. Math. Phys. **22**, 2472 (1981).

[8] W. Boenkost, Diploma thesis (in German), University of Frankfurt, 1987.

[9] R. Constantinescu, K. Klöckner, and U. Scharffenberger, "Borel summability in the disorder parameter of the averaged Green's function for Gaussian disorder," Commun. Math. Phys. **98**, 203 (1985).

[10] F. Constantinescu and U. Scharffenberger, "Borel transform, resurgence and the density of states: Lattice Schrödinger operators with exponential disorder," J. Phys. A: Math. Gen. **18**, L999 (1985).

[11] F. Constantinescu and U. Scharffenberger, "Asymptotics of the cut discontinuity and large order behavior from the instanton singularity; the case of lattice Schrödinger operators with exponential disorder," Commun. Math. Phys. **106**, 467 (1986).

[12] J. Ecalle, "Les fonctions resurgents," Publications mathematique d'Orsay No. 81-05, 06, 07, Université d'Orsay, 1981.

Characteristic functions and higher order Lagrangians

H. A. Buchdahl
Department of Physics and Theoretical Physics, Faculty of Science, The Australian National University, Canberra, 2601, Australia

(Received 26 October 1987; accepted for publication 30 December 1987)

In the context of the extension of the Hamilton–Jacobi theory to include Lagrangians involving higher derivatives the characteristic function seems not to have been considered. This omission is here rectified.

I. INTRODUCTION

A variational problem encountered frequently—in classical dynamics, for example[1]—is to extremize the functional

$$\hat{V} := \int_t^{t'} \tilde{L}\, d\tau, \tag{1.1}$$

where the "Lagrangian" $\tilde{L}$ is a given function of N functions $\xi^A(\tau)$, their first derivatives $\dot{\xi}^A(\tau)$, and τ explicitly, $A = 1,2,\ldots,N$. In other words, if R_{N+1} is an $(N+1)$-dimensional Euclidean representative space, with coordinates x^A, τ $(A = 1,\ldots,N)$, let $P(q^1,\ldots,q^N,t)$, $P'(q'_1,\ldots,q'^N,t')$ be two arbitrarily prescribed points which have $\tau = t$, $x^A = q^A := \xi^A(t)$ and $\tau = t'$, $x^A = q'^A := \xi^A(t')$, respectively. Then, exceptional circumstances apart, the set of curves joining P and P' will contain one particular curve $\mathscr{C}$ for which the value of $\hat{V}$ has an extremum; it is this curve $\mathscr{C}$, the "extremal," which is to be found. It is well known that $\mathscr{C}$ must satisfy Euler's equations

$$\dot{\pi}_A = \frac{\partial \tilde{L}}{\partial \xi^A}, \tag{1.2}$$

where

$$\pi_A := \frac{\partial \tilde{L}}{\partial \dot{\xi}^A} \tag{1.3}$$

is the "momentum" conjugate to ξ^A. Further, by familiar methods one shows that Eqs. (1.2) are equivalent to the canonical equations

$$\dot{\xi}^A = \frac{\partial \tilde{H}}{\partial \pi_A}, \quad \dot{\pi}_A = -\frac{\partial \tilde{H}}{\partial \xi^A}, \tag{1.4}$$

where the "Hamiltonian" $\tilde{H}(\pi_1,\ldots,\pi_N, \xi^1,\ldots,\xi^N,\tau)$ is

$$\tilde{H} := \sum \pi_A \dot{\xi}^A - \tilde{L} \tag{1.5}$$

expressed as a function of the ξ^A, π_B, and τ. In turn one may make a canonical transformation so chosen that the transformed Hamiltonian vanishes; then the problem of integrating (1.4) is trivial. The generator $S(\xi^1,\ldots,\xi^N,\tau)$ of such a transformation must satisfy[2] the Hamilton-Jacobi equation

$$H\left(\frac{\partial S}{\partial \xi^1},\ldots,\frac{\partial S}{\partial \xi^N},\xi^1,\ldots,\xi^N,\tau\right) + \frac{\partial S}{\partial \tau} = 0. \tag{1.6}$$

Now, the condition $\delta\hat{V} = 0$ that $\hat{V}$ be an extremum leads to a specific path of integration $\mathscr{C}$ in (1.1), connecting assigned end points P and P'. The corresponding value of $\hat{V}$ is therefore a function $V(q'^1,\ldots,q'^N,t',q^1,\ldots,q^N,t)$ of the $2N+2$ coordinates of the end points, at each of which it satisfies a Hamilton–Jacobi equation:

$$\tilde{H}\left(\frac{\partial V}{q'^1},\ldots,\frac{\partial V}{\partial q'^N},q'^1,\ldots,q'^N,t'\right) + \frac{\partial V}{\partial t'} = 0, \tag{1.7a}$$

$$\tilde{H}\left(-\frac{\partial V}{\partial q^1},\ldots,-\frac{\partial V}{\partial q^N},q^1,\ldots,q^N,t\right) - \frac{\partial V}{\partial t} = 0. \tag{1.7b}$$

To know V is to know the solution of the variational problem since

$$\frac{\partial V}{\partial q^A} = -p^A \quad (A = 1,\ldots,N) \tag{1.8}$$

are the equations of $\mathscr{C}$, P and P' being taken as the fixed initial and variable final point, respectively. Just for this reason V goes under the name "characteristic function" of the problem or of the system to which the problem refers. (In dynamics V is Hamilton's principal function,[2] in geometrical optics it is the point characteristic,[3] and in general relativity theory it is the world function.[4])

It is not mandatory that $\tilde{L}$ should contain no derivatives of ξ^A higher than the first; therefore, let it now contain derivatives up to order n. If lowercase subscripts denote the order of derivatives with respect to τ, e.g., $\xi^A{}_a := d^a\xi^A/d\tau^a$, the Lagrangian is then $\tilde{L}(\{\xi^A{}_a\},\tau)$, $A = 1,\ldots,N$, $a = 0,\ldots,n$. Euler's equations are now[5]

$$\sum_{a=0}^{n} (-1)^a \left(\frac{\partial \tilde{L}}{\partial \xi^A{}_a}\right)_a = 0, \quad (A = 1,\ldots,N) \tag{1.9}$$

[see, also, Eq. (2.5)]. While this generalization has been known for a long time, the generalization of the Hamiltonian theory, that is to say, of the canonical equations (1.4), canonical transformations, and the Hamilton–Jacobi theory, is comparatively recent.[6] The characteristic function, however, does not seem to have been considered in this context, an omission now to be rectified. Initially the argument proceeds within the confines of the special case $N = 1$. The eventual generalization to arbitrary values of N is almost trivial, but in the meantime one can make do with a less turgid notation.

In essence, in Secs. II and III the relations (1.1)–(1.7)—with $N = 1$— are generalized to Lagrangians that involve derivatives of $\xi(\tau)$ of order > 1. In Sec. IV the characteristic function is contemplated as the appropriate solution of the differential equations (3.3) satisfied by it. By way of example, the construction of V is carried out explicitly for the case of the second-order Lagrangian $\tilde{L} = \frac{1}{2}m(\xi_1{}^2 - \omega^{-2}\xi_2{}^2)$, with constant m and ω, while Sec.

 0022-2488/88/051122-05$02.50

V B deals with certain aspects of the $(n+1)$th-order Lagrangian $\tilde{L}=(\xi_{n+1})^2$. Certain general results obtained in this case are sufficient to give the explicit form of V when $n=1,2,3$. General values of n are admitted in Sec. VI and finally, Sec. VII deals with the example of a Lagrangian considered by Constantelos[6] which has $N=3$, $n=2$.

II. VARIATION OF $\hat{V}$, $N=1$

To keep the work transparent the special case $N=1$ will first be dealt with in detail. Thus

$$\hat{V}=\int_t^{t'} \tilde{L}\,(\xi_n,\xi_{n-1},\dots,\xi_0,\tau)d\tau\,, \tag{2.1}$$

with the integration extended along some curve $\mathscr{C}$ joining $P(q,t)$ and $P'(q',t')$. In place of $\mathscr{C}$ contemplate a neighboring curve $\mathscr{C}^*$ joining neighboring points $P^*(q+\delta q,t+\delta t)$ and $P'^*(q'+\delta q',t'+\delta t')$, with "neighboring" meaning that $\mathscr{C}^*$ is given by equations of the form

$$x^A=\xi^A(\tau)+\delta\xi^A(\tau)\,, \tag{2.2}$$

granted that $\delta\xi^A:=\epsilon\zeta^A(\tau)$, where ϵ is a sufficiently small positive number and ζ^A is an arbitrary, sufficiently often differentiable function of τ. The distance between P and P^* and that between P' and P'^* both go to zero with ϵ.

When, in (2.1), the integration extends from P^* to P'^* along $\mathscr{C}^*$ and terms $O(\epsilon^2)$ are rejected, the value of the integral will be $\hat{V}+\delta\hat{V}$, where

$$\begin{aligned}\delta\hat{V}&=\int_{t+\delta t}^{t'+\delta t'} \tilde{L}\,(\xi_n+\delta\xi_n,\dots,\xi_0+\delta\xi_0,\tau)d\tau\\&\quad-\int_t^{t'}\tilde{L}\,(\xi_n,\dots,\xi_0,\tau)dt\\&=\int_t^{t'}\sum_{a=0}^{n}\tilde{L}^a\,\delta\xi_a\,d\tau+L'\,\delta t'-L\,\delta t\,,\end{aligned} \tag{2.3}$$

with $\tilde{L}^a:=\partial\tilde{L}/\partial\xi_a$ and where L',L are the values of the function $\tilde{L}(\xi_n,\dots,\xi_0,\tau)$ evaluated at P' and P, respectively. The integral $\int_t^{t'}\tilde{L}^a\,\delta\xi_a\,d\tau$ may be integrated by parts a times and (2.3) then becomes

$$\begin{aligned}\delta\hat{V}&=\sum_{a=1}^{n}\Big\{\sum_{b=0}^{a-1}(-1)^b[\tilde{L}^a{}_b\,\delta\xi_{a-b-1}]_t^{t'}\\&\quad+(-1)^a\int\tilde{L}^a{}_a\,\delta\xi_0\,d\tau\Big\}\\&\quad+\int_t^{t'}\tilde{L}\,\delta\xi_0\,d\tau+L'\,\delta t'-L\,\delta t\,.\end{aligned} \tag{2.4}$$

When the end points are fixed, $\delta t'$, δt, and the integrated parts all vanish and one arrives at the familiar result that the extremal $\mathscr{E}$ must satisfy

$$\sum_{a=0}^{n}(-1)^a\tilde{L}^a{}_a=0\,. \tag{2.5}$$

Now take $\mathscr{C}$ to be $\mathscr{E}$ and concomitantly write V in place of $\hat{V}$. Then according to (2.4), if $\delta q_c:=\delta\xi_c(t')$, $\bar{\delta}q'_c:=\delta\xi_c(t')$,

$$\delta V=\Delta\Big\{\sum_{a=1}^{n}\sum_{b=0}^{a-1}(-1)^b\tilde{L}^a{}_b\,\bar{\delta}q_{a-b-1}+L\,\delta t\Big\}\,, \tag{2.6}$$

where $\Delta G:=G'-G$ denotes the difference between the values G' and G taken at the end points P' and P by any quantity $\tilde{G}$ defined along $\mathscr{E}$. Equation (2.6) may be given a more convenient form by defining the "momenta"

$$\pi^c:=\sum_{b=0}^{n-c-1}(-1)^b\tilde{L}^{b+c+1}{}_b\,,\quad(c=0,1,\dots,n-1)\,. \tag{2.7}$$

[It may be noted that formally Eq. (2.5) of the extremal is $\pi^{(-1)}=0$.] Equation (2.6) now reads as

$$\delta V=\Delta\Big(\sum_{c=0}^{n-1}p^c\,\bar{\delta}q_c+L\,\delta t\Big)\,.$$

Since

$$\bar{\delta}q_c=\delta q_c-q_{c+1}\,\delta t\,,$$

one has

$$\delta V=\Delta\Big(\sum_{c=0}^{n-1}p^c\,\delta q_c-K\,\delta t\Big)\,, \tag{2.8}$$

where K' and K are the terminal values of

$$\tilde{K}:=\sum_{c=0}^{n-1}\pi^c\,\xi_{c+1}-\tilde{L}(\xi_n,\dots,\xi_0,\tau)\,. \tag{2.9}$$

III. THE EQUATIONS SATISFIED BY $V(N=1)$

The right-hand side (rhs) of (2.9) is a function of the $n+1$ functions $\xi_b\,(b=0,\dots,n)$, the n momenta $\pi^c\,(c=0,\dots,n-1)$, and τ. There is thus no momentum π^n conjugate to ξ_n. However, ξ_n is redundant in the sense that it may be eliminated in favor of π^{n-1} in view of (2.7); choosing $c=n-1$ in this, one has

$$\pi^{n-1}=\tilde{L}^n(\xi_n,\dots,\xi_0,\tau)\,, \tag{3.1}$$

which may be resolved for ξ_n in terms of π^{n-1}, $\xi_{n-1},\dots,\xi_0,\tau$. Eliminating ξ_n from $\tilde{K}$ in this way, it becomes a function $\tilde{H}$ of $\pi^{n-1},\dots,\pi^0,\xi_{n-1},\dots,\xi_0,\tau$. From (2.8) one now reads off the relations

$$p'^c=\frac{\partial V}{\partial q'_c}\,, \tag{3.2a}$$

$$\frac{\partial V}{\partial t'}=-H'\,, \tag{3.2b}$$

$$p^c=-\frac{\partial V}{\partial q_c}\,, \tag{3.2c}$$

$$\frac{\partial V}{\partial t}=H\,, \tag{3.2d}$$

$c=0,1,\dots,n-1$.

The function $V(q'_{n-1},\dots,q'_0,t',q_{n-1},\dots,q_0,t)$ evidently satisfies the two simultaneous equations

$$H\Big(\frac{\partial V}{\partial q'_{n-1}},\dots,\frac{\partial V}{\partial q'_0},q'_{n-1},\dots,q'_0,t'\Big)+\frac{\partial V}{\partial t'}=0\,, \tag{3.3a}$$

$$H\Big(-\frac{\partial V}{\partial q_{n-1}},\dots,-\frac{\partial V}{\partial q_0},q_{n-1},\dots,q_0,t\Big)-\frac{\partial V}{\partial t}=0\,, \tag{3.3b}$$

Remark: The variation of $\tilde{K}$ may be simplified by means of the relation

$$\tilde{L}^a=\dot{\pi}^a+\pi^{a-1}\,, \tag{3.4}$$

which follows trivially from (2.7). This leads to the relation

$$\delta\widetilde{H}=\sum_{a=0}^{n-1}(\dot{\xi}_a\,\delta\pi^a-\dot{\pi}^a\,\delta\xi_a)+\left(\frac{\partial\widetilde{L}}{\partial\tau}\right)d\tau\,. \tag{3.5}$$

The (ξ_a,π^a) are therefore pairs of canonically conjugate variables.

IV. THE CHARACTERISTIC FUNCTION AS SOLUTION ($N=1$)

Each of Eqs. (3.3) has the form of a Hamilton–Jacobi equation. Let $V^1=S(q'_{n-1},...,q'_0,t',a^{n-1},...,a^0)$ ($a^{n-1},...,a^0=\text{const}$) be a complete integral of (3.3a). Then, by inspection, $V^2=-S(q_{n-1},...,q_0,t,a^{n-1},...,a^0)$ is a complete integral of (3.3b) and

$$V^\dagger:=V^1+V^2 \tag{4.1}$$

obviously satisfied both (3.3a) and (3.3b). Then the characteristic function V is obtained from (4.1) by adjoining to it the n equations

$$\frac{\partial V^\dagger}{\partial a^k}=0\quad(k=0,...,n-1) \tag{4.2}$$

and eliminating the a^k from $V^\dagger$ by means of these.

This prescription[7] reflects the group property of canonical transformations. For convenience, in this section only, write $a^{n-1},...,a^0=:\mathbf{a}$, and analogously for any other set of n quantities, so that, for example, $V^1=S(\mathbf{q}',t',\mathbf{a})$. V^1 is the generator of a canonical transformation which takes $\mathbf{q},\mathbf{p}$ into $\mathbf{a},\mathbf{b}$, where $b^k=-\partial S(q',t',\mathbf{a})/\partial a^k$. Likewise, V^2 takes $\mathbf{a},\mathbf{b}$ into $\mathbf{q}',\mathbf{p}'$, with $b^k=-\partial S(q,t,\mathbf{a})/\partial a^k$. The two alternative equations for b^k imply exactly the relations (4.2). On the other hand, Eq. (2.8) (with H replacing K) shows that V is the generator of a canonical transformation which transforms $\mathbf{q},\mathbf{p}$ directly into $\mathbf{q}',\mathbf{p}'$. It follows that the procedure described above, i.e., the elimination of the a^k from $V^\dagger$ by means of (4.2), results just in the characteristic function V.

Before going on to general values of N some particular examples will now be considered.

V. EXAMPLES ($N=1$)

A. $n=2$: $\widetilde{L}=\frac{1}{2}m(\xi_1{}^2-\omega^{-2}\xi_2{}^2)$, $m,\omega=\text{const}$

Here

$$\widetilde{H}=-(\omega^2/2m)(\pi^1)^2+\pi^0\xi_1-\tfrac{1}{2}m\xi_1^2\,, \tag{5.1}$$

so that Eq. (3.3a) is explicitly

$$-\frac{\omega^2}{2m}\left(\frac{\partial V}{\partial q'_1}\right)^2+q'_1\left(\frac{\partial V}{\partial q'_0}\right)-\tfrac{1}{2}mq'^2_1+\frac{\partial V}{\partial t'}=0\,. \tag{5.2}$$

Setting $V=W(q'_1)+\alpha q'_0-\beta t'$, with constant α and β, one infers easily that

$$V^\dagger=\lambda^{-1}\int_{q_1}^{q'_1}F\,dx+\alpha(q'_0-q_0)-\beta(t'-t)\,, \tag{5.3}$$

where $\lambda^2:=\omega^2/2m$ and

$$F:=(-\tfrac{1}{2}mx^2+\alpha x-\beta)^{1/2}\,.$$

There are two Eqs. (4.2):

$$\int_{q_1}^{q'_1}xF^{-1}\,dx+2\lambda(q'_0-q_0)=0\,, \tag{5.4a}$$

$$\int_{q_1}^{q'_1}F^{-1}\,dx+2\lambda(t'-t)=0\,. \tag{5.4b}$$

The elimination of α and β is facilitated by the introduction of the following abbreviations:

$$u:=q'_1,\quad v:=q_1,\quad \eta:=\alpha^2-2m\beta,\quad \chi:=q'_0-q_0\,,$$
$$\mu:=mu-\alpha,\quad \nu:=mv-\alpha,\quad T:=\omega(t'-t). \tag{5.5}$$

Thus, for example,

$$2m[F(u)]^2=\eta-\mu^2,\quad 2m[F(v)]^2=\eta-\nu^2. \tag{5.6}$$

Combining (5.4a) and (5.4b) one finds at once that

$$F(u)-F(v)=\lambda(m\chi-\alpha\omega^{-1}T). \tag{5.7}$$

Explicitly evaluating the integral in (5.4b) one has

$$\arcsin(\nu\eta^{-1/2})-\arcsin(\mu\eta^{-1/2})=T,$$

whence

$$\eta=(\mu^2+\nu^2-2\mu\nu\cos T)\csc^2 T. \tag{5.8}$$

Then Eqs. (5.6) become

$$\begin{aligned}(2m)^{1/2}F(u)&=(\nu-\mu\cos T)\csc T,\\ (2m)^{1/2}F(v)&=(\nu\cos T-\mu)\csc T,\end{aligned} \tag{5.9}$$

where ambiguities of sign have been resolved by appealing to internal consistency. Equation (5.7) now becomes an equation for α that gives

$$\alpha=-m[(u+v)(1-\cos T)-\omega\chi\sin T]/f, \tag{5.10}$$

where

$$f:=T\sin T-2(1-\cos T). \tag{5.11}$$

Next, the result of an integration by parts of the integral in (5.3) may be simplified by means of (5.4), giving

$$V=(2m\lambda)^{-1}[\mu F(u)-\nu F(v)]-(\alpha^2/2m\omega)T+\alpha\chi.$$

Using (5.9) and removing μ and ν, this becomes

$$\begin{aligned}2m\omega V=\csc T\{&-f\alpha^2-2m[(u+v)(1-\cos T)\\ &-\omega\chi\sin T]\alpha\\ &+m^2[2uv-(u^2+v^2)\cos T]\}\,.\end{aligned}$$

It remains to eliminate α by means of (5.10). Then, finally,

$$\begin{aligned}V=(m/2\omega)f^{-1}\{&(\sin T-T\cos T)(q'_1{}^2+q_1{}^2)\\ &+2(T-\sin T)q'_1q_1\\ &+2\omega(1-\cos T)(q'_1+q_1)(q'_0-q_0)\}\\ &+\omega^2\sin T(q'_0-q_0)^2\,.\end{aligned} \tag{5.12}$$

It may be noted in passing that the characteristic function which belongs to the Lagrangian $\widetilde{L}=\frac{1}{2}\xi_2{}^2$ may be obtained from the rhs of (5.12) as follows: First, reverse its sign; second, set $m=\omega^2$; and last, take the limit $\omega\to 0$. It turns out that, with $z:=t'-t$,

$$\begin{aligned}V=\;&2z^{-1}(q'_1{}^2+q'_1q_1+q_1{}^2)-6z^{-2}(q'_1+q_1)(q'_0-q_0)\\ &+6z^{-3}(q'_0-q_0)^2\,.\end{aligned} \tag{5.13}$$

The equation of the extremal whose initial data are p^1,p^0,q_1,q_0 is given by the pair $\partial V/\partial q_0=-p^0$, $\partial V/\partial q_1=-p^1$ between which q'_1 is to be eliminated.

B. Generalities concerning $\tilde{L}=\frac{1}{2}(\xi_{n+1})^2$

Even very simple Lagrangians that contain derivatives of order >2 seem to present great difficulties when it comes to finding the characteristic function by standard methods. In special cases one may, however, follow a different route. In this context we shall here merely consider a particular class of Lagrangians, namely polynomial Lagrangians, homogeneous quadratic in ξ and its derivatives. This entails the linearity of Euler's equations. In turn this implies that generically V is a quadratic form in the terminal values q'_a, q_a of $\xi_a(\tau)$ $(a=0,\ldots,n)$, granted that $\tilde{L}$ is of order $n+1$. Thus

$$V=\sum_{j=0}^{n}\sum_{k=0}^{n}\left(\frac{1}{2}A^{jk}q'_j q'_k+B^{jk}q'_j q_k+\frac{1}{2}C^{jk}q_j q_k\right), \tag{5.14}$$

with $A^{kj}=A^{jk}$, $C^{kj}=C^{jk}$. The coefficients of this quadratic form are functions of t' and t. (Strictly speaking, they should all carry an additional index $n+1$, but this may be left understood here.) Substitution of (5.14) in (3.2b) and (3.2d) then leads to a set of coupled ordinary differential equations for them which may be solved recursively.

The simplest Lagrangian of order $n+1$ is

$$\tilde{L}=\tfrac{1}{2}(\xi_{n+1})^2 \tag{5.15}$$

and the corresponding Hamiltonian is

$$\tilde{H}=\frac{1}{2}(\pi^n)^2+\sum_{c=0}^{n-1}\pi^c\xi_{c+1}\,. \tag{5.16}$$

The Hamilton–Jacobi equations (3.3) are therefore

$$\frac{1}{2}\left(\frac{\partial V}{\partial q'_n}\right)^2+\sum_{c=0}^{n-1}q'_{c+1}\frac{\partial V}{\partial q'_c}+\frac{\partial V}{\partial t'}=0, \tag{5.17a}$$

$$\frac{1}{2}\left(\frac{\partial V}{\partial q_n}\right)^2-\sum_{c=0}^{n-1}q_{c+1}\frac{\partial V}{\partial q_c}-\frac{\partial V}{\partial t}=0. \tag{5.17b}$$

In (5.14) t' and t can only occur in the combination $z:=t'-t$ [recall (5.3)] and from dimensional analysis one concludes that

$$V=\sum_{j=0}^{n}\sum_{k=0}^{n}\left(\frac{1}{2}a^{jk}q'_j q'_k+b^{jk}q'_j q_k+\frac{1}{2}c^{jk}q_j q_k\right)\times z^{-(2n-j-k+1)}. \tag{5.18}$$

Substitution in Eqs. (5.17) gives the following relations governing the constant coefficients a^{jk}, b^{jk}, c^{jk}:

$$a^{nj}a^{nk}+a^{j,k-1}+a^{j-1,k}-(2n-j-k+1)a^{jk}=0, \tag{5.19a}$$
$$a^{nj}b^{nk}+b^{j-1,k}-(2n-j-k+1)b^{jk}=0, \tag{5.19b}$$
$$b^{nj}b^{nk}-(2n-j-k+1)c^{jk}=0, \tag{5.19c}$$
$$b^{jn}b^{kn}-(2n-j-k+1)a^{jk}=0, \tag{5.19d}$$
$$b^{jn}c^{nk}-b^{kj-1}-(2n-j-k+1)b^{jk}=0, \tag{5.19e}$$
$$c^{nj}c^{nk}-c^{j,k-1}-c^{j-1,k}-(2n-j-k+1)c^{jk}=0. \tag{5.19f}$$

(Any coefficient with an index whose value <0 is to be taken as zero.) Under the interchange of end points, $q'_k\leftrightarrow q_k$, $z\to -z$, and $V\to -V$. It thus follows from (5.18) that

$$c^{jk}=(-1)^{j+k}a^{jk}, \tag{5.20a}$$
$$b^{jk}=(-1)^{j+k}b^{kj}. \tag{5.20b}$$

As a consequence the three relations (5.19c), (5.19e), and (5.19f) have become redundant. It is convenient to label the remaining equations (5.19a), (5.19b), and (5.19d) I^{jk}, J^{jk}, K^{jk}, respectively.

I have not been able to find the general solution of these equations. Therefore, proceeding step by step, K^{00} and I^{00} jointly give $b^{0n}=ja^{n0}$, where $j^2=1$. Then K^{n0} and K^{nn} yield $b^{nn}=(n+1)j$ and $a^{nn}=(n+1)^2$ in turn, and now I^{nn} gives $a^{n,n-1}=-\frac{1}{2}n(n+1)^2(n+2)$. Continuing in this fashion, including only coefficients that have $j,k\geqslant n-2$, I find that

$$\begin{aligned}
&a^{nn}=(n+1)^2,\quad a^{n,n-1}=-\tfrac{1}{2}n(n+1)^2(n+2),\quad a^{n,n-2}=\tfrac{1}{6}(n-1)n(n+1)^2(n+2)(n+3),\\
&a^{n-1,n-1}=\tfrac{1}{3}n^2(n+1)^2(n+2)^2,\quad a^{n-1,n-2}=-\tfrac{1}{8}(n-1)n^2(n+1)^2(n+2)^2(n+3),\\
&a^{n-2,n-2}=\tfrac{1}{20}(n-1)^2n^2(n+1)^2(n+2)^2(n+3)^2,\\
&b^{nn}=j(n+1),\quad b^{n,n-1}=jn(n+1)(n+2),\quad b^{n,n-2}=\tfrac{1}{2}j(n-1)n(n+1)(n+2)(n+3),\\
&b^{n-1,n-1}=-jn(n+1)(n+2)(n^2+2n-1),\\
&b^{n-1,n-2}=-\tfrac{1}{2}j(n-1)n(n+1)(n+2)(n+3)(n^2+2n-2),\\
&b^{n-2,n-2}=\tfrac{1}{4}j(n-1)n(n+1)(n+2)(n+3)(n^4+4n^3-3n^2-14n+16).
\end{aligned} \tag{5.21}$$

Using only Eqs. (5.19), the actual value of j cannot be found: The presence of j merely reflects the invariance of these equations under the simultaneous sign reversal of all the b^{jk}. From the relation $b^{0n}=ja^{n0}$ already derived, one finds, using (5.20), that $b^{n0}=(-1)^nja^{n0}$ and $c^{n0}=(-1)^na^{n0}$. Consequently, the terms of V governed by a^{n0} are $a^{n0}(q'_n+jq_n)(q'_0+(-1)^njq_0)$. However, just as t' and t must occur together in the combination $t'-t$, so can q'_0,q_0 occur here only in the combination q'_0-q_0 [cf. (5.3)]. It follows at once that

$$j = (-1)^n. \tag{5.22}$$

Equation (5.21) is now sufficiently complete to give V explicitly for $n = 0,1$, and 2. When $n = 0$,

$$V_{n=0} = \tfrac{1}{2}(q_0 - q_0)^2 z^{-1}.$$

When $n = 1$, (5.21) reproduces (5.13) exactly. When $n = 2$,

$$\begin{aligned} V_{n=2} = {} & \tfrac{1}{2}(3q'^2_2 - 2q'_2 q_2 + 3q_2^2)z^{-1} \\ & - 12(3q'_2 q'_1 + 2q'_2 q_1 - 2q_2 q'_1 - 3q_2 q_1)z^{-2} \\ & + 60(q'_2 - q_2)(q'_0 - q_0)z^{-3} \\ & + 24(4q'^2_1 + 7q'_1 q_1 + 4q_1^2)z^{-3} \\ & - 360(q'_1 + q_1)(q'_0 - q_0)z^{-4} \\ & + 360(q'_0 - q_0)^2 z^{-5}. \end{aligned} \tag{5.23}$$

VI. GENERAL VALUES OF *N*

It remains to extend the preceding work to general values of N. The Lagrangian is now a function of N functions $\xi^A{}_0(\tau)$ $(A = 1,...,N)$; their derivatives of order a, $\xi^A{}_a(\tau)$ $(a = 0,...,n)$; and τ explicitly. Therefore, in place of (2.1),

$$\widehat{V} = \int_t^{t'} \widetilde{L}\,(\xi^N{}_n,...,\xi^N{}_0;...;\xi^1{}_n,...,\xi^1{}_0;\tau)\,d\tau. \tag{6.1}$$

If $\widetilde{L}_A{}^a := \partial\widetilde{L}/\partial\xi^A{}_a$, one has the generalization

$$\delta\widehat{V} = \int_t^{t'} \sum_{A=1}^{N} \sum_{a=0}^{n} \widetilde{L}_A{}^a\,\delta\xi^A{}_a\,d\tau + L'\,\delta t' - L\,\delta t. \tag{6.2}$$

Formally, (6.2) differs from (2.3) only through the appearance of a second summation, namely, that over the additional index A. The generalization of the equations following (2.3) likewise involves no more than a second summation. The "momenta" are now

$$\pi_A{}^a := \sum_{b=0}^{n-a-1} (-1)^b L_A{}^{a+b+1b}. \tag{6.3}$$

(If one formally assigns the value -1 to a one has the equations $\pi_A^{(-1)} = 0$ of the extremals.) Then

$$\delta V = \Delta\left(\sum_{A=1}^{N}\sum_{a=0}^{n-1} p_A{}^a\,\delta q^A{}_a - K\,\delta t\right), \tag{6.4}$$

where

$$\widetilde{K} := \sum_{A=1}^{N}\sum_{a=0}^{n-1} \pi_A{}^a\,\xi^A{}_{a+1} - \widetilde{L}. \tag{6.5}$$

From (6.3) one has, in particular, the N relations $\pi_A{}^{n-1} = \widetilde{L}_A{}^n$. Provided $\det(\partial^2\widetilde{L}/\partial\xi^A{}_{n-1}\,\partial\xi^B{}_{n-1}) \neq 0$, these may be solved for the $\xi^A{}_n$ and the $\xi^A{}_n$ may thus be eliminated from (6.5) in favor of the $\pi_B{}^{n-1}$. $\widetilde{K}$ is now a function $\widetilde{H}$ of the $2nN+1$ variables $\pi_A{}^a\xi^A{}_a,\tau$ $(A = 1,...,N,\ \ a = 0,...,n-1)$. [The $(\pi_A{}^a\xi^A{}_a)$ are canonically conjugate pairs of variables; cf. the end of Sec. III.] From (6.4) one now reads off the equations

$$p'_A{}^a = \frac{\partial V}{\partial q'^A{}_a}, \tag{6.6a}$$

$$\frac{\partial V}{\partial t'} + H' = 0, \tag{6.6b}$$

$$p_A{}^a = -\frac{\partial V}{\partial q^A{}_a}, \tag{6.6c}$$

$$\frac{\partial V}{\partial t} - H = 0. \tag{6.6d}$$

(6.6b) and (6.6d) are the two Hamilton–Jacobi equations which the characteristic function $V(\{q'^A{}_a\},\{q^A{}_a\},t',t)$ must satisfy $(A = 1,...,N, a = 0,...,n-1)$, granted that in H' and H the $p'_A{}^a$ and $p_A{}^a$ are replaced by the derivatives of V according to (6.6a) and (6.6c). To find V, i.e., the characteristic function, from (6.6b) and (6.6d), one proceeds essentially as in Sec. IV.

VII. EXAMPLE (N=3, n=2)

To illustrate the preceding results we choose the Lagrangian

$$L = \frac{1}{2} m \sum_{A=1}^{3} [(\xi^A{}_1)^2 - \omega^{-2}(\xi^A{}_2)^2] \tag{7.1}$$

considered by Constantelos.[6] Regard $\xi^A{}_0,\xi^A{}_1$, $\pi_A{}^0,\pi_A{}^1$ as the components of Euclidean three-vectors $\boldsymbol{\xi}$, $\dot{\boldsymbol{\xi}}$, $\boldsymbol{\pi}^0$, $\boldsymbol{\pi}^1$, respectively, together with the corresponding vectors $\mathbf{q}$, $\dot{\mathbf{q}}$, $\mathbf{p}^0$, $\mathbf{p}^1$ and their primed counterparts. Then

$$\widetilde{H} = -\lambda^2|\boldsymbol{\pi}^1|^2 + \boldsymbol{\pi}^0\cdot\dot{\boldsymbol{\xi}} - \tfrac{1}{2}m|\dot{\boldsymbol{\xi}}|^2, \tag{7.2}$$

where $\widetilde{H}$ is the sum of three time-independent Hamiltonians, each of the form (5.1). It is not difficult to convince oneself that as a consequence the characteristic function V associated with the Lagrangian (7.1) can be read from (5.12): One only needs to replace in the factors multiplying the various functions of T each of the terms bilinear in q'_0, q_0, q'_1, q_1 by the corresponding scalar product. Thus

$$\begin{aligned} V = {} & (m/2m)f^{-1}\{(\sin T - T\cos T)(|\dot{\mathbf{q}}'|^2 + |\dot{\mathbf{q}}|^2) \\ & + 2(T - \sin T)\dot{\mathbf{q}}'\cdot\dot{\mathbf{q}} \\ & - 2\omega(1 - \cos T)(\mathbf{q}' + \mathbf{q})\cdot(\mathbf{q}' - \mathbf{q}) \\ & + \omega^2 \sin T\,|\mathbf{q}' - \mathbf{q}|^2\}. \end{aligned} \tag{7.3}$$

[1] V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978).
[2] E. T. Whittaker, *Analytical Mechanics* (Cambridge U.P., Cambridge, 1952), 4th ed.
[3] H. A. Buchdahl, *Introduction to Hamiltonian Optics* (Cambridge U.P., Cambridge, 1970).
[4] J. L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1960).
[5] L. E. Elsgolc, *Calculus of Variations* (Pergamon, Oxford, 1961).
[6] G. C. Constantelos, Nuovo Cimento B **84**, 91 (1984).
[7] C. Lanczos, *The Variational Principles of Mechanics* (Univ. of Toronto, Toronto, 1970), 4th ed.

Notes on a class of homogeneous space-times

M. O. Calvão, M. J. Rebouças, and A. F. F. Teixeira
Centro Brasileiro de Pesquisas Físicas, Rua Xavier Sigaud 150, Urca, 22290, Rio de Janeiro, RJ, Brazil

W. M. Silva, Jr.
Universidade Federal Fluminense, Instituto de Física, Outeiro de S. J. Batista S/No-Centro, 24100, Niterói, RJ, Brazil and Universidade Federal do Rio de Janeiro, Instituto de Física, Bloco A-S/320-Centro de Tecnologia, Ilha do Fundão, 21944, Rio de Janeiro, RJ, Brazil

(Received 16 October 1987; accepted for publication 13 January 1988)

The breakdown of causality in homogeneous Gödel-type space-time manifolds is examined. An extension of Rebouças–Tiomno (RT) and Accioly–Gonçalves studies is made. The existence of noncausal curves is also investigated under two different conditions on the energy-momentum tensor. An integral representation of the infinitesimal generators of isometries is obtained, extending previous works on the RT geometry.

I. INTRODUCTION

Ever since the foundations of general relativity were laid, there have been investigations on the potentialities of this theory, particularly as concerns its consistency with Mach's principle, the solutions of its field equations, causality conditions, and the like.

A number of familiar space-times make it clear that general relativity, as it is normally formulated, does not exclude the violation of causality in large scale, despite its local Lorentzian character. The Gödel model[1] is perhaps the best known example of a cosmological solution of Einstein's field equations in which causality may be violated.

The existence of closed timelike curves in all homogeneous Gödel-type Riemannian manifolds was examined in a recent paper by Rebouças and Tiomno.[2] They have shown that the causality main features of these space-times depend upon two independent parameters, m and Ω. For $0 \leqslant m^2 < 4\Omega^2$ there exists only one noncausal region whereas for $m^2 < 0$ there are an infinite number of alternating causal and noncausal regions. They have also found that for $m^2 \geqslant 4\Omega^2$ there is no closed timelike curve of the Gödel type.

Very recently the homogeneous Gödel-type space-times have been discussed in the framework of the higher derivative gravity (HDG) theory by Accioly and Gonçalves.[3] Regarding the $m^2 = 4\Omega^2$ metric, they have shown that it is also a solution of the HDG theory. Then, by using results from Ref. 2, they go as far as to state that they have "succeeded in finding completely causal solutions."

However, not only Rebouças and Tiomno but also Accioly and Gonçalves have restricted their study to the section $t = z = \text{const}$ (cylindrical coordinates) of the Gödel-type space-time manifolds. In other words, they have only examined the breakdown of causality of the type that occurs in the Gödel universe, leaving open the question of whether or not there is a distinct type of violation of causality.

In this paper we extend these investigations by examining the existence of all types of closed timelike curves in the homogeneous Gödel-type Riemannian space-times. We also examine whether or not they are stably causal. Moreover, by using the Newman–Penrose[4] null tetrad techniques we discuss the breakdown of causality in these space-times in connection with two different algebraic Segrè types of the energy-momentum tensor. There emerges from our results that among the new Rebouças-Tiomno[2] (RT) solutions, the special one with $m^2 = 4\Omega^2$ is the unique globally causal Gödel-type solution with an algebraic Segrè characteristic $[(1,11)1]$. Nevertheless, it is not stably causal. We also find an integral representation of the infinitesimal generators of isometries for the special RT space-time, extending previous works on this subject matter.[2,5–9]

II. MAIN RESULTS AND DISCUSSIONS

In dealing with the causal structure of space-time manifolds, the most general and powerful approach is undoubtedly the one based upon topological techniques.[10–13] As a matter of fact, several other distinct problems in general relativity also require these techniques to a greater or lesser extent.[11–13] However, most proofs employing the topological approach tend to be rather long and to have a somewhat technical character.

In what follows we shall adopt instead a simpler procedure, already used by Penrose,[14] Maitra,[15] and Ozsváth and Schücking[16] among others. We should perhaps state from the outset that our treatment does hold as long as the manifold is homeomorphic to $\mathbb{R}_4$. This is not as strong a constraint as it might appear at first sight: the Kasner and the Gödel space-times, the plane wave solutions, certain Weyl solutions, the open Friedmann models, the solutions representing collapsing spherical dust clouds as well as the Minkowski space-time, just to mention a few, all have the same underlying manifold $\mathbb{R}_4$.[11]

It is known that all Gödel-type Riemannian manifolds homogeneous in space and time (hereafter called ST homogeneous) can be put into the form[2]

$$ds^2 = [dt + H(r)d\phi]^2 - D^2(r)d\phi^2 - dr^2 - dz^2, \tag{2.1}$$

where the functions $H(r)$ and $D(r)$ are given by

$$\text{(i)}\quad H = (2\Omega/\mu^2)[1 - \cos(\mu r)], \qquad D = (1/\mu)\sin(\mu r), \tag{2.2}$$

when $\mu^2 = -m^2 = \text{const} > 0$;

 0022-2488/88/051127-03$02.50

(ii) $H=\Omega r^2, \quad D=r,$ (2.3)

whenever $m=0$;

(iii) $H=(2\Omega/m^2)[\cosh(mr)-1],$

$D=(1/m)\sinh(mr),$ (2.4)

if $m^2=\text{const}>0$. In all cases Ω is a constant.

The existence of closed timelike curves of the Gödel type, i.e., the circles defined by $t,z,r=\text{const}$, depends on the behavior of the function

$$G(r)=D^2(r)-H^2(r). \quad (2.5)$$

Indeed, if $G(r)$ becomes negative for a certain range of values of r ($r_1<r<r_2$, say), Gödel's circles, $t,z,r=\text{const}$, are closed timelike curves.

It is, therefore, not difficult to show that there are closed timelike curves in the above classes (i) and (ii). As for the third class ($m^2>0$) Rebouças and Tiomno[2] have found that for $m^2<4\Omega^2$ there is one noncausal region. They have also shown that for $m^2\geqslant4\Omega^2$ there is no violation of causality of the Gödel type[1] (Gödel's circles).

Since the presence of a single closed timelike curve is sufficient to ensure the breakdown of causality, the existence of noncausal curves, other than Gödel's circles, in the RT class of homogeneous space-times $m^2\geqslant4\Omega^2$ remains to be examined.

We shall now prove that there are no closed timelike curves in the $m^2\geqslant4\Omega^2$ space-time manifolds. To this end, we first introduce new coordinates t', x, and y defined by[2]

$$\tan[\phi/2+(m^2/4\Omega)(t'-t)]=e^{-mr}\tan(\phi/2),$$
$$e^{mx}=\cosh(mr)+\sinh(mr)\cos\phi, \quad (2.6)$$
$$mye^{mx}=\sinh(mr)\sin\phi,$$

and rewrite the line element for the hyperbolic family of space-times (2.1) and (2.4) in the form

$$ds^2=(dt'+(2\Omega/m)e^{mx}\,dy)^2-e^{2mx}\,dy^2-dx^2-dz^2. \quad (2.7)$$

where $-\infty<t',x,y,z<+\infty$, rendering explicit that the manifold has been endowed with the $\mathbb{R}_4$ topology. Now following Maitra's reasoning,[15] suppose there is a closed timelike curve in this family of space-times, represented by the parametric equations

$$x^\mu=x^\mu(\lambda). \quad (2.8)$$

Along the curve, each function $x^\mu(\lambda)$ then either is a constant or has one or more extrema. Therefore, in both cases there is a point P where $x^0(\lambda)$ satisfies

$$\dot{x}^0|_P=0. \quad (2.9)$$

Here the overdot denotes $d/d\lambda$. Now from (2.7) the vector field $dx^\mu/d\lambda$ tangent to the curve is such that

$$\dot{x}^\mu\dot{x}_\mu|_P=(4\Omega^2/m^2-1)\dot{y}^2-\dot{x}^2-\dot{z}^2. \quad (2.10)$$

For $m^2\geqslant4\Omega^2$ this equation implies that

$$\dot{x}^\mu\dot{x}_\mu|_P\leqslant0, \quad (2.11)$$

in contradiction with our initial hypothesis. Thus there are no closed timelike curves for the $m^2\geqslant4\Omega^2$ family of Riemannian space-time manifolds.

It should be stressed that although the above procedure has been applied to a special class of space-times, it does hold for all space-time manifolds homeomorphic to $\mathbb{R}_4$, which can thus be covered by a single coordinate patch.

The existence of closed timelike curves in a given space-time is by far the most unequivocal manifestation of its causal anomalies. Nevertheless, there are space-times that have no closed timelike curves, and yet an arbitrarily small perturbation of their metrics would produce causality violation—they are "on the verge" of displaying breakdown of causality. These space-times are said to violate stable causality.[10,11,13] From the general relativity point of view, a stably causal space-time is generally agreed to have a satisfactory causal behavior.

We shall now prove that the $m^2>4\Omega^2$ family of Gödel-type Riemannian space-times is stably causal. Indeed, for the function $f=t'$, from Eq. (2.7) one has

$$g^{\mu\nu}f_{,\mu}f_{,\nu}=1-4\Omega^2/m^2. \quad (2.12)$$

implying that the gradient $f_{,\mu}$ is strictly timelike provided $m^2>4\Omega^2$. In this case, therefore, f is a global time function. The existence of such a function is a remarkable feature of the $m^2>4\Omega^2$ class of space-times. It implies that all space-time manifolds of this family are stably causal.[17] Particularly, they have neither timelike nor null closed curves.

From the above procedure, the $m^2=4\Omega^2$ space-time is not stably causal. Actually this was already expected, since for the $m^2<4\Omega^2$ class there exists violation of causality.

Before proceeding to the discussion of the connection between breakdown of causality and the types of the Ricci spinor, let us state the problem and fix our notation. The algebraic classification of the symmetric second-order Ricci tensor (or spinor) in general relativity is an eigenvalue problem with an underlying four-dimensional space-time endowed with a metric of signature -2. This problem gives rise to the Segrè types, which can be specified in terms of the Segrè characteristics. It turns out that only the types [1,111] and [2,11] and their specializations are consistent with both the dominant energy condition and the local Lorentzian character of general relativity. In referring to the Segrè types we use a notation where the individual digits inside square brackets are related to the multiplicity of the corresponding eigenvalue, equal eigenvalues are enclosed in parentheses, and the first digit corresponds with a timelike or null eigenvector and is separated from the spacelike ones by a comma.

The necessary condition for the Ricci spinor $\phi_{A\dot{B}}$ to be of Segrè type [1,(111)] or Segrè type [(1,11)1] is that the Plebański spinor vanishes identically,[18] viz.,

$$\chi_{ABCD}=\tfrac{1}{4}\phi_{(AB}{}^{\dot{E}\dot{F}}\phi_{CD)\dot{E}\dot{F}}\equiv0. \quad (2.13)$$

However, for the Gödel-type metric (2.1) one has $\phi_{0\dot{2}}=0$. Thus Eq. (2.13) implies

$$\phi_{0\dot{1}}=\phi_{1\dot{2}}=\tfrac{1}{8}(H'/D)'=0. \quad (2.14)$$

where a prime means a derivative with respect to r.

Now for $\phi_{A\dot{B}}$ to be of type [1,(111)] and [(1,11)1] one has to demand, respectively, that[19]

$$\phi_{0\dot{0}}=\phi_{2\dot{2}}=2\phi_{1\dot{1}} \quad (2.15)$$

and

$$\phi_{0\dot{0}} = \phi_{2\dot{2}} = -2\phi_{1\dot{1}}. \tag{2.16}$$

A rather lengthy but straightforward calculation,[20] checked by using the set of algebraic computer programs CLASSI,[21] gives the values of $\phi_{A\dot{B}}$, which together with Eqs. (2.15) and (2.16) furnish, respectively,

$$m^2 = 2\Omega^2 \tag{2.17}$$

and

$$m^2 = 4\Omega^2, \tag{2.18}$$

where $m^2 = D''/D$ and $2\Omega = H'/D$.

Equation (2.17) defines nothing but the Gödel model, which is known to violate the causality principle. We have, therefore, shown that all Gödel-type Riemannian space-time manifolds of algebraic perfect fluid Segrè type have closed timelike curves. Similarly, Eq. (2.18) characterizes the Rebouças–Tiomno[2] metric and therefore, bearing in mind the above results, we conclude that all Gödel-type Riemannian space-time manifolds of Segrè type [(1,11)1] have no closed timelike curves. Furthermore, they are not stably causal.

In Euclidean geometry the metric relations are unaffected by translations and rotations. Real gravitational fields do not usually have such a high degree of symmetry. Nevertheless, they often admit some continuous group of transformations preserving their structure. A *conformal motion*, for example, preserves the metric up to a factor whereas a *motion* (or isometry) preserves the metric itself. The group of isometries of a space-time manifold is, undoubtedly, the most important group of symmetries as far as metric theories are concerned.

In the remainder of this paper we shall be concerned with the isometric transformations of the RT space-time, whose line element can be brought into the form

$$ds^2 = \cosh^2(r)dr^2 - dr^2 - \sinh^2(r)d\phi^2 - dz^2 \tag{2.19}$$

by a trivial coordinate transformation. For the sake of simplicity we have set $c = \Omega = 1$. It has been shown by Teixeira *et al.*[5] that besides the trivial Killing vector fields ∂_t, ∂_ϕ, and ∂_z, the metric (2.19) admits four additional Killing vector fields, which can be written in a collective notation as[22]

$$K(\epsilon,\theta) = \sin(\phi + \epsilon t + \theta)[\coth(r)\partial_\phi + \epsilon\tanh(r)\partial_t] - \cos(\phi + \epsilon t + \theta)\partial_r. \tag{2.20}$$

where $\epsilon = \pm 1$ and $\theta = 0$ or $\pi/2$.

Even when the set of infinitesimal generators of isometries of a certain space-time is known, finding an explicit finite transformation mapping the manifold onto itself can, in many cases, be a rather difficult task to perform. Nevertheless, we did succeed in obtaining the integral representation for the Killing vector fields (2.20) of RT space-time. They are collectively given by

$$\begin{aligned} t &= t' - \epsilon\cot^{-1}[\coth(a)\coth(r')\csc(\phi' + \epsilon t' + \theta) + \cot(\phi' + \epsilon t' + \theta)], \\ \cosh(2r) &= \cosh(2r')\cosh(2a) + \sinh(2r')\sinh(2a)\cos(\phi' + \epsilon t' + \theta), \\ \cot(\phi + \epsilon t + \theta) &= \cot(\phi' + \epsilon t' + \theta)\cosh(2a) + \sinh(2a)\coth(2r')\csc(\phi' + \epsilon t' + \theta), \end{aligned} \tag{2.21}$$

where a is an arbitrary real constant. The last two equations make it clear what kind of transformation is involved: a translation by a distance a in a hyperbolic plane orthogonal to the z axis, in a direction which makes an angle $\epsilon\Omega t + \theta$ with the $\phi = 0$ axis. The first equation (2.21) gives the time transformation necessary to fix each isometry.

ACKNOWLEDGMENTS

We are grateful to Professor M. A. H. MacCallum and Dr. N. A. Lemos for raising questions that motivated the present work. One of us (W.M.S., Jr.) also acknowledges Dr. M. M. Som for useful discussions. Finally, we thank an anonymous referee for suggestions that prompted us to clarify several statements on causal properties.

[1] K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).
[2] M. J. Rebouças and J. Tiomno, Phys. Rev. D **28**, 1251 (1983).
[3] A. J. Accioly and A. T. Gonçalves, J. Math. Phys. **28**, 1547 (1987).
[4] E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962); **4**, 998(E) (1962).
[5] A. F. F. Teixeira, M. J. Rebouças, and J. E. Åman, Phys. Rev. D **32**, 3309 (1985).
[6] M. J. Rebouças, J. E. Åman, and A. F. F. Teixeira, J. Math. Phys. **27**, 1370 (1986).
[7] M. J. Rebouças and A. F. F. Teixeira, Phys. Rev. D **34**, 2985 (1986).
[8] J. D. Oliveira, A. F. F. Teixeira, and J. Tiomno, Phys. Rev. D **34**, 3661 (1986).
[9] M. J. Rebouças and J. E. Åman, J. Math. Phys. **28**, 888 (1987).
[10] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U. P., Cambridge, 1973).
[11] R. Geroch and G. T. Horowitz, "Global structure of space-time," in *General Relativity, An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U. P., Cambridge, 1979).
[12] B. O'Neill, *Semi-Riemannian Geometry* (Academic, New York 1983).
[13] R. M. Wald, *General Relativity* (University of Chicago, Chicago, 1984).
[14] R. Penrose, Rev. Mod. Phys. **37**, 215 (1965).
[15] S. C. Maitra, J. Math. Phys. **7**, 1025 (1966).
[16] I. Ozsváth and E. L. Schücking, Ann. Phys. (NY) **55**, 166 (1969).
[17] Actually, a space-time admits a global time function if and only if it is stably causal. See S. W. Hawking, Proc. R. Soc. London Ser. A **308**, 433 (1969).
[18] G. C. Joly and M. A. H. MacCallum, "Computer-aided classification of the Ricci tensor in general relativity," Class. Quantum Gravit. (to be published).
[19] C. B. G. McIntosh, J. M. Foyster, and A. W.-C. Lun, J. Math. Phys. **22**, 2620 (1981).
[20] For the sake of brevity we state our results without going into details of calculations that may be easily worked out from Ref. 6.
[21] J. E. Åman, "A Manual for CLASSI. Classification programs for geometries in general relativity," USIP Report No. 77/14, 1977.
[22] As a matter of fact, Ref. 5 exhibits a different infinitesimal representation of these seven isometries.

Absence of phase transitions for continuum models of dimension $d > 1$

David Klein
Department of Mathematics, Univerisity of Southern California, Los Angeles, California 90089
Wei-Shih Yang
Department of Mathematics, University of California at Los Angeles, California 90024

(Received 11 June 1987; accepted for publication 20 January 1988)

Examples of grand canonical continuum models are given in $\mathbb{R}^d$, or a suitable subset of $\mathbb{R}^d$, for which no multiple phases exist.

I. INTRODUCTION AND MAIN RESULT

The majority of results on classical continuum models of statistical mechanics establish, in some way, uniqueness of phase and/or decay of correlations in the high temperature or low activity regions of the (grand canonical) thermodynamic parameters. In constrast to the theory of phase transitions for lattice models, few inroads have been made into the description of multiple phase regions for continuum models. Positive results on the existence of multiple phases in some continuum models, however, can be found in Bricmont, Kuroda, and Lebowitz,[1] Israel[2] (see his Appendix B), and the references contained in those works.

In this paper we give examples of continuum models in $\mathbb{R}^d$, or a suitable Borel subset R of $\mathbb{R}^d$, for which no multiple phases, and hence no phase transitions, exist. The examples, given in Sec. II, are constructed from the hypotheses of Theorem 1, stated below, which generalizes a result of one of the authors in Ref. 3. The proof of Theorem 1 is given in Sec. III and is based on the method of Dobrushin.[4]

We consider finite range, superstable, many body interactions V of the form

$$V(x) = \sum_{N=1}^{|x|} \sum_{\substack{y \subset x \\ |y| = N}} \varphi_N(y), \tag{1.1}$$

where x is a finite configuration of cardinality $|x|$, and $\inf_{N,y} \varphi_N(y) > -\infty$. Define $V(\phi) = 0$ for the empty configuration ϕ. We do not necessarily assume translation invariance for V.

For configurations $x \subset \Lambda$ and $s \cap \Lambda^c$, denote by $V(x|s)$, as in Refs. 3, 5, and 6, the energy of x assuming the external configuration $s \cap \Lambda^c$. The finite volume conditional Gibbs measure for volume Λ, external configuration $s \cap \Lambda^c$, fugacity z, and inverse temperature β is given by

$$\mu_\Lambda(dx|s) = \{\exp[-\beta V(x|s)]/Z_\Lambda(s,\beta,z)\} \nu_\Lambda(dx), \tag{1.2}$$

where

$$\nu_\Lambda(dx) = \sum_{n=0}^{\infty} \frac{z^n}{n!} d^n x$$

is an unnormalized Poisson measure in Λ and $Z_\Lambda(s,\beta,z)$ makes $\mu_\Lambda(dx|s)$ a probability measure. A Gibbs state is a probability measure on the set of all configurations in R whose conditional probabilities are determined by (1.2). For details, see Refs. 5 and 7–9.

We can now state the main result.

Theorem 1: Let the interaction V satisfy the restrictions given for Eq. (1.1). Suppose there exists an increasing sequence of bounded Borel sets $\{\Lambda_k\}$ whose union is $R \subseteq \mathbb{R}^d$ such that

(1) $\varphi_N(y) = 0$, for any $N \geqslant 2$ *and* $y = (y_1,...,y_N)$ such that $y \cap \Lambda_k \neq \phi \neq y \cap \Lambda^c_{k+1}$ for some k;

(2) $\sum_k \left[\sup_s Z_{A_k}(s,\beta,z)\right]^{-1}$ diverges, where $A_k = \Lambda_k / \Lambda_{k-1}$.

Then the Gibbs state for V,β,z is unique.

Remark 1.1: Existence of the Gibbs state under the hypotheses of Theorem 1 can be established via the methods of Refs. 7 and 9.

For the examples of the next section, we will assume that the interaction given by (1.1) satisifies one of the following two conditions.

Condition A: $\varphi_N(y) \geqslant 0$, for all N and all $y = (y_1,...,y_N)$.

For Condition B, let

$$\psi_N(y_1,...,y_N) = \begin{cases} \infty, & \text{if } \max_{ij} \|y_i - y_j\| < r_0, \\ 0, & \text{otherwise,} \end{cases}$$

where $\|\cdot\|$ denotes a Euclidean norm on $\mathbb{R}^d$, r_0 is less than the range r of the interaction V, and N is an integer greater than 1.

Condition B: There exists an integer $N \geqslant 2$ and an interaction V' satisfying the restrictions given for Eq. (1.1) such that

$$V(x) = V'(x) + \sum_{\substack{y \subset x \\ |y| = N}} \psi_N(y),$$

for any finite configuration x.

Remark 1.2: If V satisfies Condition B with $N = 2$, then V is a hard-core interaction in the usual sense. However, if V satisfies Condition B for large N, small r_0, and V' has no hard-core restrictions, then the behavior of particles with interaction V should be almost the same as the behavior of particles with no hard-core restrictions (and with interaction V'). Condition A or Condition B implies the following inequalities:

$$V(x_1,...,x_n|s) > -Bn, \tag{1.3}$$

for some $B > 0$, all boundary configurations s, and all configurations $(x_1,...,x_n)$. If $A_k \subset R$ has positive Lebesgue measure $|A_k|$, then

 0022-2488/88/051130-04$02.50

$$V(x|s) > -c|A_k|, \tag{1.4}$$

for some $c>0$, all finite $x \subset A_k$, and any boundary configuration $s \cap A_k^c$.

II. EXAMPLES

A. Wedge in $\mathbb{R}^d$

In this example R is a wedge in $\mathbf{R}^d$ properly containing the cylinder

$$\{(x_1,\dots,x_d)\in\mathbf{R}^d:\ -\infty < x_1 < \infty,\ x_2^2 + \cdots + x_d^2 \leqslant c\},$$

for some $c>0$. We assume R can be expressed as

$$R = \bigcup_{k=1}^{\infty} \Lambda_k,$$

where, for k sufficiently large, Λ_k is contained in a right cylinder, centered at the origin, of length $2rk$ along the x_1 axis. For any $c>0$, the cross-sectional area a_k satisfies

$$a_k < c \ln k, \tag{2.1}$$

for all k sufficiently large, depending on c. Let V be translation invariant within R with range equal to r so that condition 1 of Theorem 1 is fulfilled. If V satisfies Condition A or Condition B, then, by (2.1) and Remark 1.2,

$$\sup_s Z_{A_k}(s,\beta,z) \leqslant \int \exp[\beta c|A_k|]\,\nu_{A_k}(dx)$$
$$= \exp[(\beta c + z)|A_k|] < k.$$

Hence condition 2 of Theorem 1 is satisfied for all β and z and there is no phase transition.

Remark 2.1: Chayes and Chayes consider wedges in $\mathbf{Z}^d$ in Ref. 10. They prove the existence of a nontrivial low temperature critical point for the Ising magnet and bond percolation if the cross-sectional area of the wedge diverges logarithmically with its width. Example A above is a complementary result and together with Ref. 10 suggests that a phase transition in the continuum wedge R will occur if a_k does not satisfy (2.1) for all $c>0$ and all large k. We note that theorems for a variety of lattice models, analogous to Theorem 1, can be shown to hold using the techniques of Sec. III.

B. Hard rods in $\mathbb{R}^d$

Here $R = \mathbf{R}^d$ and the interaction V satisfies Condition A and models a system of hard rods all of equal length 1 and parallel to the x axis. The cross-sectional diameter of the rods is not translation invariant and decreases with the distance of the rod to the x axis in such a way as to fulfill the conditions given below.

Let Λ_k be the right circular cylinder of length $2k$, centered at the origin of $\mathbf{R}^d$ and symmetric about the x axis, with cross-sectional area $a_k = \ln(\ln k)$ for k sufficiently large.

We assume that V satisfies condition 1 of Theorem 1 (i.e., a rod whose center lies in Λ_k cannot "feel" a rod whose center lies in Λ_{k+1}^c). It is easily verified that

$$\sum_k Z_{A_k}(\phi,\beta,z)^{-1} \geqslant \sum_k \exp[-z|A_k|]$$
$$= \sum_k \exp[-z(ka_k - (k-1)a_{k-1})]$$

diverges for all β and z. Since

$$\sup_s Z_{A_k}(s,\beta,z) = Z_{A_k}(\phi,\beta,z)$$

condition 2 of Theorem 1 is satisfied and there is no phase transition.

C. Increasing external force

Let $R = \mathbb{R}^d$ and let

$$\sum_{N=2}^{|x|} \sum_{\substack{y\subset x\\ |y|=N}} \varphi_N(y)$$

satisfy Condition A or Condition B, the restrictions for Eq. (1.1), and assume each $\varphi_N(\cdot)$ is translation invariant. Let V be defined by

$$V(x) = \sum_{N=1}^{|x|} \sum_{\substack{y\subset x\\ |x|=N}} \varphi_N(x),$$

where, for $y\in\mathbf{R}^d$,

$$\varphi_1(y) \geqslant [(d-1)/\beta]\ln\|y\| \equiv \bar{\varphi}_1(y), \tag{2.2}$$

for $\|y\|$ sufficiently large. Let Λ_k be the hypersphere of radius $r(k+1)$, where $r>0$ is a constant. If the range of V is less than or equal to r, then condition 1 of Theorem 1 is fulfilled. To see that condition 2 of Theorem 1 also holds, observe that

$$\sup_s Z_{A_k}(s,\beta,z) \leqslant \int \exp\Big[\beta B|x| - \beta \sum_{i=1}^{|x|} \varphi_1(x_i)\Big]\nu_{A_k}(dx)\,. \tag{2.3}$$

Then from (2.2) and the definition of Λ_k,

$$\sup_s Z_{A_k}(s,\beta,z)$$
$$\leqslant \sum_{n=0}^{\infty} \exp[\beta Bn - \beta n\bar{\varphi}_1(rk)]\,\frac{z^n}{n!}\,|A_k|^n$$
$$= \exp[z|A_k|\exp(\beta B)\exp(-(d-1)\ln(rk))],$$

for k sufficiently large. Since $|A_k| \leqslant ck^{d-1}$ for some constant $c>0$,

$$\sup_s Z_{A_k}(s,\beta,z) \leqslant \exp[zck^{d-1}\exp(\beta B)\,(rk)^{-d+1}]$$
$$= \exp[cz\exp(\beta B)/r^{d-1}]$$

and $\Sigma_k\,[\sup_x Z_{A_k}(s,\beta,z)]^{-1}$ clearly diverges. Thus if (2.2) holds, the Gibbs measure is unique for all z. If, for any β,

$$\varphi_1(y) \geqslant [(d-1)/\beta]\ln\|y\|,$$

for all $\|y\|$ sufficiently large, then there is no phase transition.

Remark 2.2: The external force $\varphi_1(\cdot)$ satisfying (2.2) has the effect of lowering the critical temperature, for the translation invariant system described in this subsection, to a value smaller than $1/\beta$.

D. Radially decreasing range of interaction

Let $R = \mathbf{R}^d$ and let Λ_n be the hypersphere of volume

$$\sum_{k=1}^{n} ca_k$$

centered at the origin of $\mathbf{R}^d$, where $c>0$. Let V satisfy Condition A and condition 1 of Theorem 1. If

$$\sum_k \exp[-za_k]$$

diverges all of $z>0$, then condition 2 of Theorem 1 holds and there is no phase transition. The case

$$a_k \equiv 1 \tag{2.4}$$

was considered in Ref. 3. For application to the one-dimensional case where (2.4) is natural see Refs. 3, 4, and 6.

III. PROOF OF THEOREM 1

The proof of Theorem 1 uses the following notation and background information.

Definition 3.1: Let $X(\Lambda)$ denote the set of configurations in the Borel set $\Lambda \subset \mathbb{R}^d$. For an interaction V, bounded Borel set $\Lambda \subset \widetilde{\Lambda}$ with positive Lebesgue measure, and boundary condition $s \cap \widetilde{\Lambda}$, the finite volume Gibbs density $r_{\widetilde{\Lambda}}^{\Lambda}(x|s)$ is the density for $\mu_{\widetilde{\Lambda}}(dx|s)$, restricted to $X(\Lambda)$, with respect to $\nu_\Lambda(dx)$. From (1.2),

$$r_{\widetilde{\Lambda}}^{\Lambda}(x|s) = \int_{X(\widetilde{\Lambda}\setminus\Lambda)} \frac{\exp[-\beta V(x\cup y|s)]}{Z_{\widetilde{\Lambda}}(s,\beta,z)} \nu_{\widetilde{\Lambda}|\Lambda}(dy).$$

Definition 3.2: If μ_1 and μ_2 are two probability measures with densities r_1 and r_2 with respect to the finite measure ν, define

$$\rho(r_1,r_2) = \frac{1}{2}\int |r_1(x)-r_2(x)|\nu(dx)$$
$$= 1 - \int \min[r_1(x),r_2(x)]\nu(dx).$$

Note that $\rho(r_1,r_2)$ is the variation distance between μ_1 and μ_2.

An application of Dobrushin's lemma (see Ref. 4, Lemma 1) assuming condition 1 of Theorem 1 shows

$$\rho(r_{\Lambda_n}^{\Lambda_k}(\ |s),r_{\Lambda_n}^{\Lambda_k}(\ |t)) \leqslant \alpha\rho(r_{\Lambda_n}^{\Lambda_{k+1}}(\ |s),r_{\Lambda_n}^{\Lambda_{k+1}}(\ |t)), \tag{3.1}$$

for any configurations $s\cap\Lambda_n^c$, $t\cap\Lambda_n^c$, where $n>k+3$ and

$$\alpha = 1 - \inf_{s,t}\int_{X(\Lambda_k)} \min\left[\frac{\exp[-\beta V(x|s)]}{Z_{\Lambda_k}(s,\beta,z)}, \frac{\exp[-\beta V(x|t)]}{Z_{\Lambda_k}(t,\beta,z)}\right]\nu_{\Lambda_k}(dx). \tag{3.2}$$

For details see Refs. 3 and 4.

Proof of Theorem 1: As in Refs. 3, 4, and 6, it suffices to show that

$$\lim_{n\to\infty}\sup_{s,t}\rho(r_{\Lambda_n}^{\Lambda_{k_0}}(\ |s),r_{\Lambda_n}^{\Lambda_{k_0}}(\ |t)) = 0, \tag{3.3}$$

for any fixed k_0 sufficiently large.

For any configuration x and $\Lambda\subset R$, let $x_\Lambda \equiv x\cap\Lambda$. In particular ϕ_{A_k} denotes the empty configuration in A_k. For any three external configurations $s,t,u\in X(\Lambda_k^c)$ and $y\in X(\Lambda_{k-1})$,

$$\min\left[\frac{\exp[-\beta V(\phi_{A_k}\cup y|s)]}{Z_{\Lambda_k}(s,\beta,z)}, \frac{\exp[-\beta V(\phi_{A_k}\cup y|t)]}{Z_{\Lambda_k}(t,\beta,z)}\right] \geqslant \frac{\exp[-\beta V(\phi_{A_k}\cup y|u)]}{\sup_s Z_{\Lambda_k}(s,\beta,z)} = \inf_s \frac{\exp[-\beta V(\phi_{A_k}\cup y|s)]}{Z_{\Lambda_k}(s,\beta,z)}. \tag{3.4}$$

Integrating both sides of (3.4) gives

$$\int_{\{\phi_{A_k}\}\cup X(\Lambda_{k-1})} \min\left[\frac{\exp[-\beta V(x|s)]}{Z_{\Lambda_k}(s,\beta,z)}, \frac{\exp[-\beta V(x|t)]}{Z_{\Lambda_k}(t,\beta,z)}\right]\nu_{\Lambda_k}(dx)$$
$$= \int_{X(\Lambda_{k-1})} \min\left[\frac{\exp[-\beta V(\phi_{A_k}\cup y|s)]}{Z_{\Lambda_k}(s,\beta,z)}, \frac{\exp[-\beta V(\phi_{A_k}\cup y|t)]}{Z_{\Lambda_k}(t,\beta,z)}\right]\nu_{\Lambda_{k-1}}(dy)$$
$$\geqslant \inf_s \frac{1}{Z_{\Lambda_k}(s,\beta,z)}\int_{X(\Lambda_{k-1})} \exp[-\beta V(\phi_{A_k}\cup y|u)]\nu_{\Lambda_{k-1}}(dy)$$
$$= \inf_s \int_{X(\Lambda_{k-1})} \frac{\exp[-\beta V(\phi_{A_k}\cup y|s)]}{Z_{\Lambda_k}(s,\beta,z)}\nu_{\Lambda_{k-1}}(dy). \tag{3.5}$$

The right side of (3.5) may be rewritten as

$$\inf_s \int_{X(\Lambda_{k-1})}\int_{X(A_k)} \frac{\exp[-\beta V(x|y\cup s)]}{Z_{A_k}(y\cup s,\beta,z)} \frac{\exp[-\beta V(\phi_{A_k}\cup y|s)]}{Z_{\Lambda_k}(s,\beta,z)} \nu_{A_k}(dx)\nu_{\Lambda_{k-1}}(dy)$$
$$= \inf_s \int_{X(\Lambda_k)} \frac{1}{Z_{A_k}(y_{\Lambda_{k-1}}\cup s,\beta,z)}\mu_{\Lambda_k}(dy|s),$$

which is bounded below by

$$\left[\sup_s Z_{A_k}(s,\beta,z)\right]^{-1}.$$

Hence

$$\inf_{s,t}\int_{X(\Lambda_k)} \min\left[\frac{\exp[-\beta V(x|s)]}{Z_{\Lambda_k}(s,\beta,z)}, \frac{\exp[-\beta V(x|t)]}{Z_{\Lambda_k}(t,\beta,z)}\right]\nu_{\Lambda_k}(dx) \geqslant \frac{1}{\sup_s Z_{A_k}(s,\beta,z)}. \tag{3.6}$$

Combining (3.1), (3.2), and (3.6) gives

$$\rho(r_{\Lambda_n}^{\Lambda_{k_0}}(\ |s), r_{\Lambda_n}^{\Lambda_{k_0}}(\ |t)) < (1 - h_{k_0})\rho(r_{\Lambda_n}^{\Lambda_{k_0+1}}(\ |s), r_{\Lambda_n}^{\Lambda_{k_0+1}}(\ |t)), \tag{3.7}$$

for any s, t, and k_0, where

$$h_k = \left(\sup_s Z_{A_k}(s,\beta,z)\right)^{-1}.$$

Applying (3.7) inductively shows

$$\sup_{s,t} \rho(r_{\Lambda_n}^{\Lambda_{k_0}}(\ |s), r_{\Lambda_n}^{\Lambda_{k_0}}(\ |t))$$

$$< \left[\prod_{k=k_0}^{k_0+m} (1 - h_k)\right] \sup_{s,t} \rho(r_{\Lambda_n}^{\Lambda_{k_0+m}}(\ |s), r_{\Lambda_n}^{\Lambda_{k_0+m}}(\ |t)),$$

for $m < n + 3$. Equation (3.3) now follows from the fact that

$$\prod_{k=k_0}^{\infty} (1 - h_k) = 0,$$

when $\Sigma_{k=k_0}^{\infty} h_k$ diverges, which is condition 2 of Theorem 1. This completes the proof.

ACKNOWLEDGMENT

One of us (W.-S. Y.) was partially supported by National Science Foundation Grant No. DMS 8601800.

[1] J. Bricmont, K. Kuroda, and J. Lebowitz, "First order phase transitions in lattice and continuum systems: Extensions of Pirogov–Siani theory," Commun. Math. Phys. **101**, 501 (1985).

[2] R. B. Israel, *Convexity in the Theory of Lattice Gases* (Princeton U. P., Princeton, NJ, 1979).

[3] D. Klein, "A class of continuum models with no phase transitions," J. Math. Phys. **27**, 2872 (1986).

[4] R. L. Dobrushin, "Conditions for the absence of phase transitions in one-dimensional classical systems," Math. Sb. **93**, 29 (1974).

[5] D. Klein, "Dobrushin uniqueness techniques and the decay of correlations in continuum statistical mechanics," Commun. Math. Phys. **86**, 227 (1982).

[6] D. Klein, "Uniqueness of continuum one-dimensional Gibbs states for slowly decaying interaction," J. Stat. Phys. **43**, 207 (1986).

[7] D. Ruelle, "Superstable interactions in classical statistical mechanics," Commun. Math. Phys. **18**, 127 (1970).

[8] R. A. Minlos, "Lectures on statistical physics," Russ. Math. Surveys **23**, 137 (1968) (English translation).

[9] C. Preston, "Random fields," in *Lecture Notes in Mathematics*, Vol. 534 (Springer, Berlin, 1976).

[10] J. T. Chayes and L. Chayes, "Critical points and intermediate phases on wedges of $\mathbf{Z}^d$," J. Phys. A: Math. Gen. **19**, 3033 (1986).

Geometric Lagrangian approach to first-order systems and applications

José F. Cariñena, Carlos López, and Manuel F. Rañada
Departamento de Física Teórica, Facultad de Ciencias, Universidad de Zaragoza, 50009 Zaragoza, Spain

(Received 4 August 1987; accepted for publication 18 November 1987)

The geometric theory of presymplectic systems is developed to study both the Lagrangian and Hamiltonian formulations of a system described by a Lagrangian linear in the velocities. The results are used to study some related problems for second-order differential equations and regular Lagrangians.

I. INTRODUCTION

First-order differential equations arise frequently in the mathematical description of the deterministic time evolution of certain physical systems. In relation to classical mechanics the fundamental role is played by second-order differential equations (hereafter shortened to SODE), although the Legendre transformation FL associated with a regular Lagrangian L allows us to describe the time evolution of such systems by means of Hamilton's equations, which are first-order systems. However, it is well-known that some important Lagrange equations, such as the relativistic Dirac equation, are first-order systems. Another situation where one deals with first-order Lagrange equations, when considering a SODE, is obtained by a doubling of the dimensions of the configuration space with the introduction of new independent variables corresponding to the old velocities; i.e., the old velocity space becomes the configuration space, and the integral curves of the SODE vector field are determined by a set of first-order differential equations.

In recent years much attention has been paid to the geometrical approach to mechanics. So the Lagrangian and the Hamiltonian formulations of mechanics are considered, at least when L is regular, as being particular cases of a more general structure: the Hamiltonian dynamical systems. The dynamics is then given by a vector field Γ_L (resp. Γ_H) defined by $i(\Gamma_L)\omega_L = dE_L$ [resp. $i(\Gamma_H)\omega_0 = dH$]. But the point we want to stress now is that, if L is regular or, in other words, if ω_L is symplectic, then the (uniquely defined) solution Γ_L is a SODE; consequently, the projection on the configuration space Q of the integral curves of $\Gamma_L \in \mathscr{X}(TQ)$ gives a set of curves on which the Euler–Lagrange equations hold. On the contrary, if L is singular, then the previous assertion is not true, and the Euler–Lagrange equations defined by L may play no dynamical role but will appear as associated with SODE conditions.

First-order differential equations very often arise not only in various branches of theoretical physics, but in other fields such as biology dynamics. The knowledge of the existence of a variational formulation may be useful in the simplification of some of these problems, and it is now well-known that a Lagrangian giving rise to such a system of first-order differential equations is linear in the velocities. The singular character of such Lagrangians is referred to explicitly in some recent books of classical mechanics (see, for example, Sudarshan and Mukunda[1]). Nevertheless, the studies of these Lagrangians[2] are usually done without using the tools of modern differential geometry.[3]

We aim in this paper to apply the tools of the geometry of presymplectic systems to deal with the Lagrangians linear in velocities. So in Sec. II we develop the geometric approach for a system described by a Lagrangian $L = \hat{\mu} + \tilde{h}$, with $\mu \in \Lambda^1(Q)$, $h \in C^\infty(Q)$, and study the different possibilities according to the rank of $d\mu$ and the relation between μ and h. In this approach we will make use of the difference between dynamical and SODE constraints, and these methods are applied to some examples. In Sec. III we establish the relation of the Lagrangian to the Hamiltonian formalism. The explicit relations between the constraint functions arising in both formulations are given using the K operator introduced by Batlle *et al.*,[4] the geometric version of which was recently discussed.[5] The theory so developed is not only an academic subject, but the results obtained for Lagrangians linear in the velocities will be used in Sec. IV, where the inverse problem is reviewed from a new perspective. We deal in a more geometric way with an approach developed by several authors, where second-order systems are studied through first-order systems obtained by doubling the number of degrees of freedom as dynamical systems defined in $T(TQ)$. The existence of a Lagrangian $\mathbb{L} \in C^\infty[T(TQ)]$ is always asserted, and an answer to the inverse problem for a SODE is obtained with this new optics. When studying the case in which it is defined by a Lagrangian $L \in C^\infty(TQ)$, the well-known Helmholtz conditions are recovered in the geometric version of Crampin.[6] Finally, in Sec. V, we give a new application of the theory. We prove, given the Hamiltonian dynamical system (TQ, ω_L, dE_L) defined by a regular Lagrangian, that this system's infinitesimal symmetries, not given by complete lifts of vector fields in the base and, consequently, not seen as gauge symmetries of L (recently characterized by Marmo and Mukunda[7]), are such that their complete lifts are gauge symmetries for $\mathbb{L}$ and so are more easily exhibited in this approach. As a simple example, the two-dimensional isotropic harmonic oscillator is given.

II. DYNAMICS: THE CONSTRAINT ALGORITHM

We consider a Lagrangian dynamical system with an n-dimensional configuration manifold Q, and whose Lagrangian function $L \in C^\infty(TQ)$ is assumed to take the form

$$L = \hat{\mu} + \tilde{h}, \tag{2.1}$$

where $\tilde{h}$ is the pullback through the tangent projection τ_Q: $TQ \to Q$ of a function $h \in C^\infty(Q)$, and $\hat{\mu} \in C^\infty(TQ)$ denotes a

 0022-2488/88/051134-09$02.50

function linear on the fibers associated with the one-form $\mu\in\Lambda^1(Q)$, defined by

$$\hat{\mu}(q,v) = \langle \mu_q, v\rangle.$$

In local coordinates $\{q^i,v^i\}$ $(i = 1,...,n)$ adapted to the bundle, the Lagrangian takes the form

$$L = m_j(q)v^j + h(q) \tag{2.2}$$

when the one-form $\mu\in\Lambda^1(Q)$ is given by $\mu = m_j(q)dq^j$, so that (2.1) represents a Lagrangian linear in the velocities and, therefore, singular. The coefficients m_j are often written in the form[2] $m_j = m_{ji}(q)q^i$, the functions m_{ji} being in most cases independent of q.

The dynamics associated with a Lagrangian $L\in C^\infty(TQ)$ is given, in the geometric approach, by the vector fields $X\in\mathscr{X}(TQ)$, solutions of the equation[3]

$$i(X)\omega_L - dE_L = 0\,, \tag{2.3}$$

where $E_L = \Delta(L) - L$ is the energy function and $\omega_L = -d(dL\circ S)$ is the presymplectic form associated with the Lagrangian. Here Δ denotes the Liouville vector field and S is the vertical endomorphism.[8]

The expression of ω_L in local coordinates is

$$\omega_L = \frac{\partial^2 L}{\partial v^i\,\partial q^j}\,dq^i\wedge dq^j + \frac{\partial^2 L}{\partial v^i\,\partial v^j}\,dq^i\wedge dv^j\,, \tag{2.4a}$$

and the matrix representation is given by

$$\omega_L = \begin{bmatrix} A & -W \\ W & 0\end{bmatrix}, \tag{2.4b}$$

where

$$A_{ij} = \frac{\partial^2 L}{\partial q^i\,\partial v^j} - \frac{\partial^2 L}{\partial v^i\,\partial q^j}\,, \tag{2.5a}$$

$$W_{ij} = \frac{\partial^2 L}{\partial v^i\,\partial v^j}\,. \tag{2.5b}$$

When ω_L is symplectic, it follows that there is a unique solution $\Gamma_L\in\mathscr{X}(TQ)$ of Eq. (2.3) that is a second-order differential equation (SODE) vector field, i.e., verifying[8] $S(\Gamma_L) = \Delta$, but for singular Lagrangians this is not the case, and Eq. (2.3) is not equivalent to the Euler–Lagrange equations unless this SODE condition $S(X) = \Delta$ is additionally imposed. Two other properties of singular dynamical systems should be noted.

(i) Nonexistence of a global dynamics: The dynamics is restricted to a submanifold M of TQ where Eq. (2.3) can be consistently solved. The submanifold M is determined by the so-called dynamical constraint functions obtained through a geometric algorithm.[9] When the SODE conditions are also imposed some additional constraints appear, determining a smaller submanifold N.[10,11]

(ii) Ambiguity of the dynamics: The solution of (2.3) on M or N is not unique, the ambiguity of the solution being given by $\ker\omega_L \cap TM$ and $V(\ker\omega_L)\cap TN$ on M and N, respectively, where $\ker\omega_L$ is the characteristic distribution of the Lagrangian two-form ω_L and $V(\ker\omega_L)$ denotes its vertical part.

For the singular Lagrangian (2.1) we find that $E_L = -\tilde{h}$ and $\omega_L = -\widetilde{d\mu}$, where $\widetilde{d\mu}$ denotes the pullback to TQ of the two-form $d\mu$ defined in the base space $d\mu\in\Lambda^2(Q)$. The Hessian matrix W is null in all the points of TQ and thus the distribution $\ker\omega_L$ contains at least all the vertical fields [sections of $V(TQ)$] and verifies the property

$$\dim\ker\omega_L < 2\dim[V(\ker\omega_L)]\,,$$

which is a dimensional restriction characterizing the Lagrangians of type III.[12]

A. The constraint algorithm

The first step in the constraint algorithm is to consider the submanifold M_1 of TQ determined by the constraints

$$\langle dE_L, \ker\omega_L\rangle = 0\,, \tag{2.6}$$

in which Eq. (2.3) can be solved.

We recall the general result $\langle dE_L, V(\ker\omega_L)\rangle \equiv 0$, so that Eq. (2.6) reduces in our case to

$$\langle dh, \ker d\mu\rangle = 0, \tag{2.7}$$

which give restrictions on the base Q but not on the fibers.

The systematic procedure for the search of constraint functions consists of (i) determination of the elements $\xi\in\mathbf{R}^n$ such that $W\xi = 0$; (ii) choice of the elements ξ in the kernel of W such that $\langle \xi', A\xi\rangle = 0$, $\forall\xi'\in\ker W$; and then (iii) the dynamical constraint functions are $\langle\xi,\alpha\rangle = 0$, with

$$\alpha_j = \frac{\partial L}{\partial q^j} - v^k\,\frac{\partial^2 L}{\partial v^j\,\partial q^k}\,,$$

for those ξ satisfying (ii).

The ξ's of (i) determine the elements

$$\xi^i\frac{\partial}{\partial v^i}\in V(\ker\omega_L)\,,$$

and those of (ii) correspond to those which are the image under S of an element of $\ker\omega_L$. Finally, the α_j are the components of the semibasic one-form $i(D)\omega_L - dE_L$, with D being an arbitrary SODE.

In the particular case we are studying, $d\mu$ is given in local coordinates by

$$d\mu = \tfrac{1}{2}A_{ij}\,dq^i\wedge dq^j \quad (i,j = 1,...,n),$$

with

$$A_{ij} = \frac{dm_j}{\partial q^i} - \frac{\partial m_i}{\partial q^j}\,.$$

The first step above is trivial, and the second amounts to looking for the null eigenvectors of the matrix A. If n^0 denotes its rank, and

$$\{Z_a = (z^i)_a\} \quad (a = 1,...,n - n^0)$$

is a basis of $\ker A$, then Eq. (2.7) turns out to be

$$\phi_a \equiv (z^i)_a\,\frac{\partial h}{\partial q^i} = 0 \quad (a = 1,...,n - n^0)\,. \tag{2.8}$$

The general solution of (2.3) on M_1 is

$$X = \{\eta^i + \lambda^a(z^i)_a\}\,\frac{\partial}{\partial q^i} + f^i\,\frac{\partial}{\partial v^i}\,, \tag{2.9}$$

with η^i being a particular solution of

$$A_{ij}\,\eta^j = \frac{\partial h}{\partial q^i}\,, \tag{2.10}$$

and λ^a, f^i being arbitrary functions on TQ.

An interesting case occurs when $(Q,d\mu)$ is a symplectic manifold (for which n must be even-dimensional) because then A_{ij} is regular, $\ker \omega_L = V(\ker \omega_L)$, and there will be no dynamical constraints. In this case the solution is globally defined,

$$X = \eta^i \frac{\partial}{\partial q^i} + f^i \frac{\partial}{\partial v^i},$$

and the η^j's are uniquely determined by

$$\eta^i = -(A^{-1})^{ij} \frac{\partial h}{\partial q^j}.$$

In the general case, secondary constraints may appear when the consistency conditions (the solution X has to be tangent to M_1, determining M_2 and so on) are taken into account. For the Lagrangian (2.1) some of the unknown functions λ^a of the expression (2.9) are determined, but all the f^i's will remain arbitrary. In fact, all the constraints restrict only the basis Q and can be obtained by applying the constraint algorithm to the generalized Hamiltonian system $(Q,d\mu, -dh)$. Denoting by Q' the final constraint submanifold of this system and by F the set of consistent solutions of (2.3) on Q', the final constraint submanifold of the system $(TQ,d\tilde{\mu}, -d\tilde{h})$ turns out to be

$$M \equiv T_Q Q' = \{v \in TQ \,|\, \tau_Q(v) \in Q'\},$$

and every consistent solution X projects pointwise onto an element of F.

B. The second-order condition

The matrix form of the dynamical equation reduces in this case to

$$\begin{bmatrix} A_{ij} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a^j \\ b^j \end{bmatrix} = -\begin{bmatrix} \partial h/\partial q^i \\ 0 \end{bmatrix},$$

where a^j and b^j are the components of the vector field X, i.e.,

$$X = a^i \frac{\partial}{\partial q^i} + b^i \frac{\partial}{\partial v^i}.$$

In order to find a solution of SODE type and, consequently, to have a solution with $a^i = v^i$, we must impose as additional constraints

$$\langle \xi, \alpha_j \rangle = 0, \quad \forall \xi \in \ker W, \tag{2.11}$$

or, in local coordinates,

$$\phi_i = A_{ij} v^j + \frac{\partial h}{\partial q^i} = 0. \tag{2.12}$$

These constraints are considered as equations of motion in the traditional formulation.[1,2] In the geometric formulation the constraints (2.8) and (2.11) are both obtained from[13]

$$\langle i(\Gamma_0)\omega_L - dE_L, V(TQ)^\perp \rangle = 0, \tag{2.13}$$

with Γ_0 being an arbitrary SODE and

$$V(TQ)^\perp = \{Z \in TTQ \,|\, \omega_L(Z,V) = 0, \forall V \in V(TQ)\}.$$

In the submanifold N_1 determined by (2.11), the general SODE solution is

$$\Gamma = v^i \frac{\partial}{\partial q^i} + f^i \frac{\partial}{\partial v^i},$$

with the f^i still being arbitrary functions.

In the particular case of $(Q,d\mu)$ being a symplectic manifold, only the primary constraints (2.12) appear, and the consistency condition will determine all the unknown functions f^i by

$$v^i \frac{\partial}{\partial q^i}\left(A_{jk} v^k + \frac{\partial h}{\partial q^j}\right) + f^i A_{ji} = 0, \tag{2.14}$$

so that the restriction of the consistent solution $\Gamma' = \Gamma_{0|N}$ is unique. In fact, if Y_0 is the unique solution of $i(Y_0)d\mu + dh = 0$, the submanifold $N_1 = N$ is $N = Y_0(Q)$, where we consider Y_0 as a map Y_0: $Q \to TQ$, and the relation $\Gamma_{0|N} = Y^c_{0|N}$ holds for the complete lift Y^c_0 of Y_0. The pullback of ω_L on N defines a symplectic structure (N,ω_N) isomorphic to $(Q,d\mu)$.

In the more general case, only n^0 of the unknown functions f^i are determined from (2.14), and the system will admit a family Γ of solutions depending of $n - n^0$ indeterminate functions, all of them tangent to M_1. In this case, as secondary constraints may appear, the process must be continued by looking for the submanifold M_2, and so on. Once the final constraint submanifold has been found, the reduction process for obtaining a symplectic manifold must be carried out.

C. Examples

Farias[14] studies the first-order equations associated with the Lagrangian

$$L = \tfrac{1}{2}(q^2 + q^3)v^1 - \tfrac{1}{2}q^1 v^2 + \tfrac{1}{2}(q^4 - q^1)v^3 - \tfrac{1}{2}q^3 v^4 + \{-q^2 q^3 - \tfrac{1}{2}(q^3)^2 - \tfrac{1}{2}(q^4)^2\}. \tag{2.15}$$

The matrix A is given by

$$A = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

and it is not singular. Thus, in this case, $\ker \omega_L$ is precisely the four-dimensional distribution composed by all the vertical vectors.

According to what we have previously studied there are no dynamical constraints and, consequently, this system admits a global dynamics. This means that there exists a family of fields Γ that are solutions of Eq. (2.3) defined in all the tangent bundles $TQ \equiv T\mathbb{R}$.[4] In coordinates, Γ takes the form

$$\Gamma = (v^i + \lambda^i)\frac{\partial}{\partial q^i} + f^i(q,v)\frac{\partial}{\partial v^i},$$

where the λ coefficients are to be determined by

$$A_{ij}(v^j + \lambda^j) + \frac{\partial_n h}{\partial q^i} = 0.$$

For (2.15) we obtain

$$\lambda^1 = -v^1 + q^3, \quad \lambda^2 = -v^2 - q^4,$$
$$\lambda^3 = -v^3 + q^4, \quad \lambda^4 = -v^4 - q^2,$$

so that Γ is given by

$$\Gamma = q^3\frac{\partial}{\partial q^1} - q^4\frac{\partial}{\partial q^2} + q^4\frac{\partial}{\partial q^3} - q^2\frac{\partial}{\partial q^2} + f^i(q,v)\frac{\partial}{\partial v^i}. \tag{2.16}$$

The dynamics is globally defined, but it depends on the functions $f^i(q,v)$ that remain totally undeterminate.

Let us suppose that we are interested in the SODE solutions. Then the dynamics will be limited to the four-dimensional submanifold $N \subset T\mathbb{R}^4$ defined by $\phi_i(q,v) = 0$ $(i = 1,...,4)$, where

$$\phi_1 = v^2 + v^3, \quad \phi_2 = v^1 - q^3,$$
$$\phi_3 = v^4 - v^1 + q^2 + q^3, \quad \phi_4 = v^3 - q^4.$$

When imposing $\mathscr{L}_\Gamma(\phi_i) = 0$ $(i = 1,...,4)$, where Γ is now of the form

$$\Gamma = v^i\frac{\partial}{\partial q^i} + f^i(q,v)\frac{\partial}{\partial v^i},$$

we will obtain for the $f^i(q,v)$ functions the values

$$f^1 = v^3, \quad f^2 = -v^4, \quad f^3 = v^4, \quad f^4 = -v^2,$$

for which finally we get

$$\Gamma' = v^i\frac{\partial}{\partial q^i} + v^3\frac{\partial}{\partial v^1} - v^4\frac{\partial}{\partial v^2} + v^4\frac{\partial}{\partial v^3} - v^2\frac{\partial}{\partial v^4}. \tag{2.17}$$

This field represents the only dynamics described by a SODE field tangent to the submanifold N.

Jakubiec[15] presents an example of finite-dimensional classical mechanics analogous to the Dirac equation. The Lagrangian under consideration is a function $L \in C^\infty(T\mathbb{C})$ given by

$$L = (i/2)(z^*\dot{z} - \dot{z}^*z) - z^*z, \tag{2.18a}$$

or, in real coordinates $z = (1/\sqrt{2})(x + iy)$,

$$L = \tfrac{1}{2}(y\dot{x} - x\dot{y} - x^2 - y^2). \tag{2.18b}$$

This is in the form $L = \hat{\mu} + \tilde{h}$, with $\mu \in \Lambda^1(\mathbb{R}^2)$ and $h \in C^\infty(\mathbb{R}^2)$ given by $\mu = \frac{1}{2}(y\,dx - x\,dy)$ and $h = -\frac{1}{2}(x^2 + y^2)$.

The two-form $d\mu$ with matrix representation

$$d\mu = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

is obviously symplectic, and thus there exists a global solution Γ defined in all $TQ \equiv T\mathbb{R}^2$. The SODE vector field solution of the Euler–Lagrange equations is given, in the submanifold $N \subset T\mathbb{R}^2$ defined by $N = (x, y, \dot{x} = y, \dot{y} = -x)$, by

$$\Gamma' = \dot{x}\frac{\partial}{\partial x} + \dot{y}\frac{\partial}{\partial y} - x\frac{\partial}{\partial \dot{x}} - y\frac{\partial}{\partial \dot{y}}, \tag{2.19}$$

which corresponds to the complete lift X^c_{IN} of the field $X \in \mathscr{X}(\mathbb{R}^2)$, $X = y\,\partial/\partial x - x\,\partial/\partial y$, defined in the base space and solution of $i(X)d\mu = dh$.

In the same way as the former system can be simplified to a Hamiltonian system $(\mathbb{R}^2, dy \wedge dx, h)$, Dirac theory admits an analogous simplification.

III. HAMILTONIAN FORMULATION

Let us now study the relationship between the Lagrangian formulation in TQ and the Hamiltonian formulation in T^*Q. In fact, the singular character of the Lagrangian (2.1) manifests itself in that the associated Legendre map $FL: TQ \to T^*Q$ is not a local diffeomorphism. Moreover, in this case the momenta in the image are given as functions of only the base coordinates,

$$p_j = m_j(q) \quad (j = 1,...,n), \tag{3.1}$$

while the n primary constraints,

$$\Phi_j(q,p) \equiv p_j - m_j(q) = 0, \tag{3.2}$$

determine a submanifold $P_1 = F_L(TQ)$ of T^*Q given by $P_1 = \mu(Q)$. (Here μ is considered as a map $\mu: Q \to T^*Q$.) Every fiber $T_qQ = \tau_Q^{-1}(q)$ is mapped onto a unique point $\mu(q) \in T^*_qQ$, and the Hamiltonian on P_1 is the function $H = -h(q)$. The Poisson brackets of the primary constraints are given by

$$\{\Phi_j, \Phi_k\} = A_{jk} \quad (j,k = 1,...,n), \tag{3.3}$$

with the matrix A defined in (2.5a). If n^0 denotes its rank, $n - n^0$ first-class constraints can be found by linear combinations of the Φ_j. Specifically, if $\{Z_a, a = 1,...,n - n^0\}$ is a basis of $\ker A$, then

$$\Phi_a = (z^j)_a\,\Phi_j \quad (a = 1,...,n - n^0) \tag{3.4}$$

are the first-class primary constraints because $\{\Phi_a, \Phi_j\}_{|P_1} = 0$. Applying the constraint algorithm to $(P_1, \omega_1 = j_1^*(\omega_0), dH)$, the same constraints as with $(Q, d\mu, -dh)$ are obtained because the map $\mu: Q \to P_1$ defines an isomorphism of presymplectic systems, i.e., $\mu^*(\omega_1) = d\mu$ and $\mu^*(H) = -h$.

When $d\mu$ is symplectic, the rank of the matrix A_{ij} is maximum, and all the primary constraints in the Hamiltonian formulation are second class so that the submanifold (P_1, ω_1) is symplectic and there are no secondary constraints.

The Gotay theory[10,11] relating the n-ary Hamiltonian constraints with the $(n-1)$-ary (dynamical) Lagrangian constraints by FL pullback has recently been expanded by Batlle *et al.*[4] to include the SODE non-FL-projectable constraint functions. Essentially, the method is based on the introduction of an operator K mapping $C^\infty(T^*Q)$ into $C^\infty(TQ)$, where the operator's expression in local coordinates is

$$K(f)_{(q,v)} = v^i\left(\frac{\partial f}{\partial q^i}\right)_{(FL(q,v))} + \left(\frac{\partial L}{\partial q^i}\right)_{(q,v)}\left(\frac{\partial f}{\partial p^i}\right)_{(FL(q,v))}. \tag{3.5}$$

The image under K of an n-ary first-class Hamiltonian constraint is an n-ary FL-projectable (i.e., dynamical) Lagrangian constraint, while the image under K of an n-ary second-class Hamiltonian constraint is an n-ary non-FL-projectable SODE condition. A geometric approach to these results can be found in Ref. 5.

In the case of Lagrangian (2.2), K is of the form

$$K = v^i\frac{\partial}{\partial q^i} + \left(v^j\frac{\partial m_j}{\partial q^i} + \frac{\partial h}{\partial q^i}\right)\frac{\partial}{\partial p^i}, \tag{3.6}$$

and relates the first-class primary Hamiltonian constraints $\Phi_a \in C^\infty(T^*Q)$ in (3.4) with the dynamical Lagrangian constraints $\phi_a \in C^\infty(TQ)$ given by (2.8), and all the primary Hamiltonian constraints $\Phi_j \in C^\infty(T^*Q)$ of (3.2) with all the

Lagrangian ones $\phi_j \in C^\infty(TQ)$ of (2.12). In coordinates we have

$$K(\Phi_j) = A_{ij}\, v^j + \frac{\partial h}{\partial q^j} = \phi_j \tag{3.7}$$

and

$$K(\Phi_a) = v^i A_{ij} (z^j)_a + \frac{\partial h}{\partial q^i}(z^i)_a = \phi_a \, . \tag{3.8}$$

On the other hand, the secondary Hamiltonian constraints Ψ_a obtained from the consistency condition take the same coordinate expression as the primary dynamical Lagrangian constraints ϕ_a, so that the additional SODE conditions are not recovered by FL pullback. In the same way, n-ary Lagrangian constraints can be recovered by applying the operator K to the n-ary Hamiltonian constraints. Note that while all the n-ary Hamiltonian constraints (for $n > 1$) depend only on base coordinates, their associated Lagrangian constraints are velocity dependent (because of the first term of K). As only a part of the velocities have been solved as functions of the q's, not all the n-ary Lagrangian constraints are weakly FL-projectable, which means that not all the n-ary Hamiltonian constraints are first class.

IV. THE INVERSE PROBLEM FOR FIRST- AND SECOND-ORDER SYSTEMS

It is well known that a system of n second-order differential equations can be replaced by one of $2n$ first-order equations by doubling the number of degrees of freedom.[16,17] In this section we will study the geometric approach to this problem and, in particular, we will focus our attention on the Lagrangian character of these equations. In this way we will relate the inverse problem in TQ to the inverse problem in $T(TQ)$ by using Lagrangians of the form (2.1).

We have seen in Sec. II that when the two-form $d\mu$ is symplectic there are no dynamical constraints, and the associated SODE conditions turn out to be the set of n first-order differential equations corresponding to the Euler–Lagrange equations obtained directly from the Lagrangian linear in the velocities.

The inverse problem for first-order systems consists of finding a Lagrangian of type (2.1) whose Euler–Lagrange equations are a given set of first-order differential equations. We will only be interested in the case of an even number of equations.

Geometrically, we can consider an even-dimensional manifold M and a vector field $X \in \mathscr{X}(M)$ given by

$$X = f^i(x)\, \frac{\partial}{\partial x^i} \, ,$$

whose integral curves obey a set of first-order differential equations

$$\dot{x}^i = f^i(x) \, . \tag{4.1}$$

Proposition 1: If a one-form $\mu \in \Lambda^1(M)$ is such that (i) $d\mu$ is symplectic, and (ii) the Lie derivative $\mathscr{L}_X \mu$ is closed, then there exists a local Lagrangian $\mathbb{L} \in C^\infty(TM)$ of type (2.1) that is a solution of the inverse problem for (4.1).

Proof: By hypothesis (ii) there exists a locally defined function g such that $\mathscr{L}_X \mu = dg$, and therefore

$$i(X) d\mu = \mathscr{L}_X \mu - d\,(i(X)\mu) = d\,(g - \mu(X)) \, . \tag{4.2}$$

The Lagrangian $\mathbb{L} = \hat{\mu} + \tilde{h}$, with $h = g - \mu(X)$, then gives a solution as a consequence of the results in the preceding section.

In local coordinates the problem consists of finding a skew-symmetric matrix A_{ij} of the form

$$A_{ij} = \frac{\partial m_j}{\partial x^i} - \frac{\partial m_i}{\partial x^j} \, ,$$

and a function $h(x)$ for which the equations

$$A_{ij} f^j = -\frac{\partial h}{\partial x^i} \tag{4.3}$$

hold. Then the Lagrangian takes the form $\mathbb{L} = m_j v^j + h$. It is remarkable that there always exists such a solution and that it is not unique. Two different solutions (A^1, h^1) and (A^2, h^2) may give rise to two different Lagrangians $\mathbb{L}_1$ and $\mathbb{L}_2$, not necessarily gauge equivalent. Remember that two Lagrangians are said to be gauge equivalent when they differ by the total time derivative of a function or, in local coordinates, by a term

$$\widehat{df} = \frac{\partial f}{\partial x^i}\, \dot{x}^i$$

with $f \in C^\infty(M)$. In geometric terms, this means that both Lagrangians differ through a basic closed one-form.

Definition 1: Two Lagrangians $\mathbb{L}_1$ and $\mathbb{L}_2$ of type (2.1) are said to be equivalent when their final constraint submanifolds and their consistent solutions on them coincide.[18]

The next proposition allows us to find a family of equivalent Lagrangians from a given particular solution $\mathbb{L}$ of the inverse problem, which does not cover all the possible equivalent Lagrangians but is more general than the gauge equivalence.

Proposition 2: For every $\beta \in \Lambda^1(M)$ such that $X \in \ker d\beta$ and $d(\mu + \beta)$ is symplectic, the Lagrangian $\mathbb{L} = \mathrm{L} + \hat{\beta}$ is equivalent to $\mathbb{L}$.

Proof: $i(X) d(\mu + \beta) = i(X) d\mu + i(X) d\beta = dh$.

Gauge-equivalent Lagrangians are a particular case of the former ones when β is chosen to be closed. In this case all the Lagrangians

$$\mathbb{L}' = v^j\left(m_j + \frac{\partial l}{\partial x^j}\right) + h(x) \, ,$$

with l an arbitrary function $l \in C^\infty(M)$, are gauge equivalent to $\mathbb{L}$.

The inverse problem of Lagrangian dynamics is to give necessary and sufficient conditions for a system of second-order differential equations to be that of the Euler–Lagrange equations of some regular Lagrangian function. Geometrically, we consider a configuration manifold Q and a SODE field $\Gamma \in \mathscr{X}(TQ)$, given by

$$\Gamma = v^i \frac{\partial}{\partial q^i} + f^i(q,v) \frac{\partial}{\partial v^i} \, .$$

On the basis of the integral curves on the projection, the set of second-order equations holds:

$$\ddot{q}^i = f^i(q, \dot{q}) \, . \tag{4.4}$$

It is not at all clear whether or not there exists a regular

Lagrangian function $L \in C^\infty(TQ)$ such that the vector field Γ is just the one determined by the equation $i(\Gamma)\omega_L = dE_L$, but apparently not. Then the inverse problem for second-order systems cannot always be solved. However, we can consider Γ as a first-order system on TQ, and then it is always possible to find a Lagrangian function $\mathbb{L} \in C^\infty[T(TQ)]$ of type (2.1) with a SODE field solution $\Gamma^c_{|N}$ on the final constraint submanifold N (which can be identified with Γ). On the other hand, for every regular Lagrangian $L \in C^\infty(TQ)$ there exists an associated Lagrangian $\mathbb{L}(L) \in C^\infty[T(TQ)]$ of type (2.1) equivalent to L in a sense to be more accurately established in the corollary of the next proposition.

Proposition 3: Let us consider an exact symplectic manifold $(M, -d\theta)$ and a Hamiltonian $H \in C^\infty(M)$, and denote Y as the vector field solution of $i(Y)d\theta = -dH$. Then the singular Lagrangian $\mathbb{L} \in C^\infty(TM)$ defined by $\mathbb{L} = \hat{\theta} + \tilde{H}$ is such that its solution vector field is $Y^c_{|N}$, where $N = Y(M)$.

Proof: It is enough to use the results of Sec. II.

Corollary 1: Given a regular Lagrangian $L \in C^\infty(TQ)$, the Lagrangian $\mathbb{L}(L) \in C^\infty[T(TQ)]$ defined by $\mathbb{L}(L) = \hat{\theta}_L - \tilde{E}_L$ has the same solution as L.

Proof: For a regular Lagrangian L, the dynamical equations of the system $(TQ, -d\theta_L, E_L)$ are equivalent to the Euler–Lagrange equations for L. The statement of Proposition 3 completes the proof.

In local coordinates,[19] $\mathbb{L}(L)$ is given by

$$\mathbb{L}(L) = \frac{\partial L}{\partial x^{i+n}}(\dot{x}^i - x^{i+n}) + L,$$

where $x^i = q^i$ and $x^{i+n} = v^i$.

Using the former results, the inverse problem for second-order systems can be attacked in two steps.

(i) Given a SODE $\Gamma \in \mathscr{X}(TQ)$, choose a Lagrangian $\mathbb{L} \in C^\infty[T(TQ)]$ of type (2.1), the solution of the inverse problem for Γ as a first-order system.

(ii) Search for a Lagrangian $\mathbb{L}'$ equivalent to $\mathbb{L}$ and generated from a regular Lagrangian $L \in C^\infty(TQ)$ as in Corollary 1.

If you find such a $\mathbb{L}'$, the Lagrangian L is a solution of the inverse problem for Γ as a second-order system.

The next theorem gives a useful characterization of Lagrangians $\mathbb{L} \in C^\infty[T(TQ)]$ of type (2.1) generated from Lagrangians on TQ. First, we include a lemma whose statement is used in the proof of the theorem.

Lemma 1: For every semibasic one-form $\mu \in \Lambda^1(TQ)$ and every SODE $D \in \mathscr{X}(TQ)$, these two identities hold:

$$S^*(\mathscr{L}_D\mu) = \mu, \quad \mu([\Delta, D]) = \mu(D),$$

where Δ is the dilation vector field.

Proof: It is based on the fact that there exists a $\nu \in \Lambda^1(TQ)$ such that $\mu = S^*(\nu)$ and the property[8] $S([D,V]) = -V$ for every vertical vector field V. Then, for every $X \in \mathscr{X}(TQ)$,

$$\begin{aligned}
&[S^*(\mathscr{L}_D\mu)](X) \\
&\quad = \mathscr{L}_D(\mu(SX)) - \mu([D, SX]) \\
&\quad = -(S^*\nu)([D, SX]) = \nu(SX) = \mu(X),
\end{aligned}$$

because SX is a vertical vector field, and

$$\mu([\Delta, D]) = (S^*\nu)([\Delta, D]) = \nu(\Delta) = \nu(SD) = \mu(D),$$

because $S([\Delta, D]) = \Delta$.

Theorem 1: Given a Lagrangian $\mathbb{L} \in C^\infty[T(TQ)]$ of type (2.1) ($\mathbb{L} = \hat{\mu} + \tilde{h}$) that is a solution of the inverse problem for a SODE $\mathbb{L} \in \mathscr{X}(TQ)$ [i.e., with $d\mu$ symplectic and

$$i(\Gamma^C)\omega_{\mathbb{L}}\,|_{\Gamma(TQ)} = dE_{\mathbb{L}}|_{\Gamma(TQ)}],$$

there exists a regular Lagrangian $L \in C^\infty(TQ)$ for which $\mathbb{L} = \mathbb{L}(L)$ if and only if the one-form μ is semibasic.

Proof: For every Lagrangian $L \in C^\infty(TQ)$, the Poincaré–Cartan form θ_L is semibasic by definition: $\theta_L = S^*(dL)$. On the other hand, consider a Lagrangian of type (2.1) ($\mathbb{L} = \hat{\mu} + \tilde{h}$), with $\mu \in \Lambda^1(TQ)$ semibasic and $d\mu$ symplectic, that is a solution of the inverse problem for Γ. As we can see from Proposition 3, the equation

$$i(\Gamma^C)\omega_{\mathbb{L}}\,|_{\Gamma(TQ)} = dE_{\mathbb{L}}\,|_{\Gamma(TQ)}$$

is equivalent to $i(\Gamma)d\mu = dh$, recalling from Sec. II that $\theta_{\mathbb{L}} = \tilde{\mu}$ and $E_{\mathbb{L}} = -\tilde{h}$. The Lagrangian $L = i(\Gamma)\mu + h$ has the properties

$$\begin{aligned}
\theta_L &= S^*(dL) = S^*(\mathscr{L}_\Gamma\mu - i(\Gamma)d\mu + dh) \\
&= S^*(\mathscr{L}_\Gamma\mu) = \mu, \\
E_L &= \mathscr{L}_\Delta L - L \\
&= \mu([\Delta,\Gamma] - \Gamma) + (\mathscr{L}_\Delta\mu)(\Gamma) + \mathscr{L}_\Delta h - h \\
&= d\mu(\Delta,\Gamma) + dh(\Delta) - h = -h,
\end{aligned}$$

so that

$$\mathbb{L}(L) = \hat{\theta}_L - \tilde{E}_L = \hat{\mu} + \tilde{h} = \mathbb{L}.$$

In local coordinates, suppose we have a particular solution of the inverse problem for

$$\Gamma = v^i\frac{\partial}{\partial q^i} + f^i(q,v)\frac{\partial}{\partial v^i}$$

as a first-order system. That is, suppose we have a $2n \times 2n$ matrix A_{ab} ($ab = 1,\ldots,2n$) and a function $h(q,v)$ such that

$$A_{ab}\,g^b = \frac{\partial h}{\partial x^a}, \tag{4.5}$$

where

$$x^i = q^i, \quad x^{i+n} = v^i,$$

and

$$g^i = v^i, \quad g^{i+n} = f^i \quad (i = 1,\ldots,n = \dim Q),$$

the matrix A_{ab} being of the form

$$A_{ab} = \frac{\partial m_b}{\partial x^a} - \frac{\partial m_a}{\partial x^b}.$$

If we restrict ourselves to search for Lagrangians gauge equivalent to $\mathbb{L} = m_a\dot{x}^a + h(x)$, then the conditions become very simple. This is because in order to find an

$$\mathbb{L}' = \left(m_a + \frac{\partial l}{\partial x^a}\right)\dot{x}^a + h,$$

with

$$m'_a\,dx^a = \left(m_a + \frac{\partial l}{\partial x^a}\right)dx^a$$

semibasic, it is necessary and sufficient that

$$\frac{\partial m_{j+n}}{\partial x^{i+n}} = \frac{\partial m_{i+n}}{\partial x^{j+n}}, \tag{4.6}$$

so that if the former conditions hold, then

$$m_{i+n} = -\frac{\partial l}{\partial x^{i+n}},$$

and the Lagrangian on TQ is given by

$$L = \left(m_i + \frac{\partial l}{\partial q^i}\right)v^i + h .$$

Of course, the conditions (4.6) are sufficient but not necessary, because not all the Lagrangians equivalent to $\mathbb{L}$ are gauge equivalent.

In the last part of this section we are going to show how the former treatment of the inverse problem reproduces the known Helmholtz conditions. In recent years, these conditions have been rewritten in a geometric language[6]; a given SODE $\Gamma\in\mathscr{X}(TQ)$ is Lagrangian if and only if there exists a two-form $\omega\in\Lambda^2(TQ)$ with the properties

(i) ω is symplectic (closed and nondegenerate), (4.7a)

(ii) $\mathscr{L}_\Gamma \omega = 0$, (4.7b)

(iii) $\omega(V_1,V_2) = 0, \quad \forall V_1,V_2$ vertical. (4.7c)

On the other hand, according to the results of this section, there exists a local Lagrangian for Γ if and only if there exists a one-form $\mu\in\Lambda^1(TQ)$ such that

(i) $d\mu$ is symplectic, (4.8a)

(ii) $\mathscr{L}_\Gamma \mu$ is closed, (4.8b)

(iii) μ is horizontal. (4.8c)

Conditions (4.8a) and (4.8b) are necessary and sufficient to find a solution $\mathbb{L}\in C^\infty[T(TQ)]$ of the inverse problem for Γ as a first-order system, while, when condition (4.8c) holds as well, the Lagrangian $\mathbb{L}$ is generated from a regular Lagrangian $L\in C^\infty(TQ)$, $\mathbb{L} = \mathbb{L}(L)$, in the form defined in Corollary 1.

Inasmuch as we are looking for the local existence of a Lagrangian, closed forms can be identified with exact ones. Then it is a straightforward matter to prove the equivalence between conditions (4.7a),(4.7b) and (4.8a),(4.8b), respectively, when $d\mu$ and ω are identified. Moreover, condition (4.8c) implies (4.7c) because, if μ is semibasic, then, for every pair of vertical vector fields V_1,V_2,

$$d\mu(V_1,V_2) = V_1(\mu(V_2)) - V_2(\mu(V_1)) - \mu([V_1,V_2]) = 0 .$$

Conversely, suppose that condition (4.7c) holds. This means that the restriction of μ to every fiber is exact, and there will exist a function F such that

$$dF\,|_{V(TQ)} = \mu\,|_{V(TQ)} .$$

The one-form $\mu' = \mu - dF$ still preserves conditions (4.8a), (4.8b) and is semibasic by definition.

V. SYMMETRIES AND CONSTANTS OF THE MOTION

Well-known to every physicist is the result of the celebrated (first) Noether theorem establishing a relationship between infinitesimal point symmetries of the regular Lagrangian L and constants of the motion described by L. But what has often been overlooked is that this correspondence works only for point transformations and is not one-to-one. The Lagrangian L permits the definition of Hamiltonian dynamical system (TQ,ω_L,E_L). Then the Hamiltonian version of Noether's theorem says that, for any one-parameter subgroup of canonical transformations generated by X, $\mathscr{L}_X\omega_L = 0$, and leaving E_L invariant, there is a constant of motion, namely the function g, that corresponds to X through the symplectic form ω_L. In order for the function g to be globally defined, the vector field has to be not only locally Hamiltonian, but globally Hamiltonian as well, for which $\mathscr{L}_X\theta_L$ must be exact. Conversely, given a constant of motion g, the vector field X corresponding to it by $i(X)\omega_L = dg$ generates a one-parameter subgroup of canonical transformations preserving E_L. The point now is that, if X corresponds to point transformations, i.e., if X is the complete lift of a vector field Y in the basis, then

$$\mathscr{L}_X\,\omega_L = \mathscr{L}_{Y^c}\,\omega_L = \omega_{Y^c(L)} = \omega_{X(L)} ,$$
$$\mathscr{L}_X\,E_L = \mathscr{L}_{Y^c}\,E_L = E_{Y^c(L)} = E_{X(L)} ,$$

and therefore the infinitesimal canonical transformations of a symmetry of E_L are simply those transforming L into a gauge-equivalent Lagrangian. In other words, there will exist a closed one-form β in Q such that

$$X(L) = \hat{\beta}.$$

If β is not only closed but exact ($\beta = df$), then

$$\mathscr{L}_X\,\theta_L = \theta_{XL} = \theta_{\hat{\beta}} = d\tilde{f},$$

and the function g is $i(X)\theta_L - \tilde{f}$.

If X is not a complete lift of a vector field in the basis, then the preceding relations do not hold, and either $\mathscr{L}_X\theta_L \neq \theta_{XL}$ or $\mathscr{L}_X E_L \neq E_{XL}$ results, or (even worse) both. Correspondingly, the symmetries of the Hamiltonian system (TQ,ω_L,dE_L) have nothing to do with $\mathscr{L}_X\theta_L = d\tilde{f}$, the traditional concept of gauge symmetry for L. In fact, if we want Noether's theorem to have a converse, we need to modify the concept of symmetry in an appropriate way as indicated by Marmo and Mukunda.[7] The idea is that we have to accept the general infinitesimal transformations

$$\delta q^i = \varepsilon\xi^i(q,v), \quad \delta v^i = \varepsilon\eta^i(q,v),$$

corresponding to a vector field

$$X = \xi^i\frac{\partial}{\partial q^i} + \eta^i\frac{\partial}{\partial v^i},$$

without the condition of η being the total time derivative of ξ. However, this condition will be obtained as a final subproduct of the theory of the symmetry. More accurately, in order for X to be an infinitesimal symmetry of the dynamical field Γ defined by L, it is necessary that $[X,\Gamma] = 0$, and this automatically implies that $\eta^i = \Gamma(\xi^i)$, this last condition corresponding to the weaker one, $S([X,\Gamma]) = 0$.

Marmo and Mukunda[7] associated any vector field $X\in\mathscr{X}(TQ)$ and every SODE D with the vector field

$$X(D) = X + S([D,X]),$$

which, in coordinates, reads

$$X(D) = \xi^i\frac{\partial}{\partial q^i} + D(\xi)\frac{\partial}{\partial v^i} .$$

They have been able to prove that the necessary and sufficient condition for $X = X(\Gamma)$ to be a symmetry of the Hamiltonian dynamical system (TQ, ω_L, dE_L) is that there exists a function $F \in C^\infty(TQ)$ such that $\mathscr{L}_{X(D)} L = \mathscr{L}_D F$ for every SODE D.

This condition characterizing the symmetries of the theory is somewhat academic because it may be difficult to check the condition for any SODE D. We must look for more practical criteria.

Our aim in this section is to show that these symmetries can be looked upon as classical Noether symmetries for the associated Lagrangian $\mathbb{L}(L) \in C^\infty[T(TQ)]$ of type (2.1).

Definition 2: A vector field $Y \in \mathscr{X}(TQ)$ is a Noether symmetry of the Lagrangian $\mathbb{L}(L) = \hat{\theta}_L - \widetilde{E}_L$ if there exists a function $F \in C^\infty(TQ)$ such that

$$\mathscr{L}_{Y^c} \mathbb{L} = \widehat{dF}, \tag{5.1}$$

where Y^C is the complete lift of Y to $T(TQ)$.

It is an easy task to prove that the associated constant of the motion is given by $i(Y^C)\theta_{\mathbb{L}} - \widetilde{F}$, which is a τ_{TQ}-projectable function.

Theorem 2: For every regular Lagrangian $L \in C^\infty(TQ)$, there is a one-to-one correspondence between constants of the motion defined by L and Noether symmetries of the associated Lagrangian $\mathbb{L}(L)$ according to Definition 2.

Proof: Let Γ be the SODE solution of $i(\Gamma)\omega_L = dE_L$. For every function $G \in C^\infty(TQ)$ such that $\mathscr{L}_\Gamma G = 0$, the vector field X defined by $i(X)\omega_L = dG$ has the property $\mathscr{L}_{X(D)} L = \mathscr{L}_D F$ for every D SODE, with $F = i(X)\theta_L - G$. Moreover, $\mathscr{L}_X \theta_L = dF$, and then

$$\mathscr{L}_{X^c} \mathbb{L} = \mathscr{L}_{X^c}(\hat{\theta}_L) - \mathscr{L}_{X^c}(\widetilde{E}_L) = (\widehat{\mathscr{L}_X \theta_L}) = \widehat{dF},$$

so that X is a Noether symmetry of L according to Definition 2.

On the other hand, if there exists a vector field $X \in \mathscr{X}(TQ)$ and a function $F \in C^\infty(TQ)$ such that $\mathscr{L}_{X^c} \mathbb{L} = \widehat{dF}$, then

$$\begin{aligned} 0 &= \mathscr{L}_{X^c}(\hat{\theta}_L - \widetilde{E}_L) - \widehat{dF} \\ &= (\widehat{\mathscr{L}_X \theta_L - dF}) - (\widetilde{\mathscr{L}_X E_L}). \end{aligned}$$

While the first term is linear in the fibers of the vector bundle $\tau_{TQ}: T(TQ) \to TQ$, the second term is the pullback of a function on the basis TQ, so that both terms must vanish separately. Developing the expressions, we arrive at

$$i(X)dE_L = 0 \quad \text{and} \quad i(X)\omega_L + d(i(X)\theta_L - F) = 0.$$

Contracting the second equation with Γ and taking into account the first one, we find that $\mathscr{L}_\Gamma(i(X)\theta_L - F) = 0$, so that $G = i(X)\theta_L - F$ is a constant of the motion.

Finally, as a demonstration of this theorem we can consider the two-dimensional harmonic oscilator with the Lagrangian $L \in T(\mathbb{R}^2)$ given by

$$L = \tfrac{1}{2}\{(v_1)^2 + (v_2)^2 - (q_1)^2 - (q_2)^2\}. \tag{5.2}$$

This system has two "hidden" symmetries corresponding to non-Noether constants of motion. Specifically, the corresponding constants of the motion are the two energies associated with each one of the two degrees of freedom:

$$E_i = \tfrac{1}{2}(v_i^2 + q_i^2) \quad (i = 1,2). \tag{5.3}$$

The corresponding vector fields $X_i \in \mathscr{X}(T\mathbb{R}^2)$ in the version of Marmo and Mukunda take the form (no summation is understood)

$$X^i = v_i \frac{\partial}{\partial q_i} - q_i \frac{\partial}{\partial v_i} \quad (i = 1,2), \tag{5.4}$$

so that the equations

$$\mathscr{L}_{X^i(D)} L = \mathscr{L}_D F_i \quad (i = 1,2) \tag{5.5}$$

hold for every SODE field D, with functions of $F_i \in C^\infty(T\mathbb{R}^2)$ given by $F_i = \frac{1}{2}(v_i^2 - q_i^2)$.

On the other hand, the associated Lagrangian of type (2.1) is in this case

$$\begin{aligned} \mathbb{L} &\equiv \mathbb{L}(L) \\ &= x_3\dot{x}_1 + x_4\dot{x}_2 - \tfrac{1}{2}\{(x_1)^2 + (x_2)^2 + (x_3)^2 + (x_4)^2\}, \end{aligned} \tag{5.6}$$

with $x_i = q_i$ and $x_{i+2} = v_i$ $(i = 1,2)$.

The vector fields

$$X^1 = x_3 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_3}$$

and

$$X^2 = x_4 \frac{\partial}{\partial x_2} - x_2 \frac{\partial}{\partial x_4}$$

are infinitesimal generators of rotations in the planes 1-3 and 2-4. They are point (Noetherian) symmetries of $\mathbb{L}$ in the sense that their complete lifts verify

$$X^{1c}\mathbb{L} = \frac{dF_1}{dt} = -x_1\dot{x}_1 + x_3\dot{x}_3, \tag{5.7a}$$

$$X^{2c}\mathbb{L} = \frac{dF_2}{dt} = -x_2\dot{x}_2 + x_4\dot{x}_4, \tag{5.7b}$$

where $F_i \in C^\infty[T(T\mathbb{R}^2)]$ are now pullback of functions defined in the base space $T\mathbb{R}^2$, and the fields $X^{ic} \in \mathscr{X}[T(T\mathbb{R}^2)]$ are given by

$$X^{1c} = x_3 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_3} + \dot{x}_3 \frac{\partial}{\partial \dot{x}_1} - \dot{x}_1 \frac{\partial}{\partial \dot{x}_3}, \tag{5.8a}$$

$$X^{2c} = x_4 \frac{\partial}{\partial x_2} - x_2 \frac{\partial}{\partial x_4} + \dot{x}_4 \frac{\partial}{\partial \dot{x}_2} - \dot{x}_2 \frac{\partial}{\partial \dot{x}_4}. \tag{5.8b}$$

The associated constants of the motion,

$$G_i = i(X^{ic})\theta_{\mathbb{L}} - F_i \quad (i = 1,2),$$

are projectable functions on the functions E_1 and E_2 of $T\mathbb{R}^2$.

ACKNOWLEDGMENTS

One of the authors (C. L.) thanks Diputación General de Aragón for a grant. Partial financial support by CAICYT is also acknowledged.

[1] E. C. Sudarshan and N. Mukunda, *Classical Mechanics: A Modern Perspective* (Wiley, New York, 1974).

[2] W. Scherer, "Identical Lagrangian and Hamiltonian constraint algorithms for first order systems," University of Rochester preprint UR-986, 1987.

[3] R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Reading, MA, 1978), 2nd ed.

[4] C. Batlle, J. Gomis, J. M. Pons, and N. Román-Roy, J. Math. Phys. **27**, 2953 (1986).

[5] J. F. Cariñena and C. López, Lett. Math. Phys. **14**, 203 (1987).

[6]M. Crampin, J. Phys. A **14**, 2567 (1981).
[7]G. Marmo and N. Mukunda, Nuovo Cimento B **92**, 1 (1986).
[8]M. Crampin, J. Phys. A **16**, 3755 (1983).
[9]M. J. Gotay, J. M. Nester, and G. Hinds, J. Math. Phys. **19**, 2388 (1978).
[10]M. J. Gotay and J. M. Nester, Ann. Inst. H. Poincaré A **30**, 129 (1979).
[11]M. J. Gotay and J. M. Nester, Ann. Inst. H. Poincaré A **32**, 1 (1980).
[12]F. Cantrijn, J. C. Cariñena, M. Crampin, and L. A. Ibort, J. Geom. Phys. **3**, 353 (1986).
[13]J. F. Cariñena, C. López, and N. Román-Roy, "Geometric study of the connection between the Lagrangian and Hamiltonian Constraints," Universidad de Zaragoza preprint, 1987.
[14]J. R. Farias, Hadronic J. **8**, 93 (1985).
[15]A. Jakubiec, Lett. Math. Phys. **9**, 171 (1985).
[16]S. Hojman and L. F. Urrutia, J. Math. Phys. **22**, 1896 (1981).
[17]S. Hojman and J. Gomez, J. Math. Phys. **25**, 1776 (1984).
[18]J. F. Cariñena and L. A. Ibort, J. Phys. A **18**, 3335 (1985).
[19]N. P. Konopleva and V. N. Popov, *Gauge Fields* (Harwood Academic, 1981).

Origin of the Lagrangian constraints and their relation with the Hamiltonian formulation

José F. Cariñena and Carlos López
Departamento de Física Teórica, Universidad de Zaragoza, 50009 Zaragoza, Spain

Narciso Román-Roy
Departamento de Física E.T.S.E.I.B., Universidad Politécnica de Catalunya, 08028 Barcelona, Spain

(Received 11 June 1987; accepted for publication 22 December 1987)

The theory of presymplectic systems is used for the study of mechanical systems described by singular Lagrangians in order to clarify the geometric meaning of the Euler–Lagrange equations for such systems. The two different types of primary constraint functions arising in the Lagrangian formulation are analyzed by means of the relation between the image under the vertical endomorphism of the kernel of the presymplectic form ω_L and its vertical part. The connection with the Hamiltonian Dirac theory is also studied and the theory is illustrated with several examples.

I. INTRODUCTION

Most textbooks nowadays (e.g., Refs. 1–3) divide their contents into three main chapters: Newtonian, Lagrangian, and Hamiltonian mechanics. The main guide for Newtonian mechanics is the determinism principle, according to which the knowledge of the positions and velocities of the points of a mechanical system at a fixed time determines their future positions and velocities. The idea is then to use second-order differential equations in normal form $\ddot{\mathbf{x}} = \mathbf{f}(\mathbf{x},\dot{\mathbf{x}})$, because the theorem of existence and uniqueness of the solution of such systems fits well with the determinism principle.

The introduction of Lagrangian mechanics is based on the observation that, at least for conservative systems, the equations of motion may also be seen as the Euler equations determining the curve solution of a variational problem; the action $\int L\,dt$, with $L = T - V$, is extremal for the actual path when compared with other fixed end point paths. There are very many interesting advantages in the use of the Hamilton principle because the extremal principle does not depend on the set of coordinates, and it incorporates the holonomic constraints through a particular choice of coordinates, leading in this way to a considerable simplification.

The problem of what kind of second-order differential equations can be obtained as the Euler–Lagrange equations of a Lagrangian function is known as the "inverse problem." It has received much attention[4–8] till recent years, when geometric conditions[9] were given to replace the well-known Helmholtz conditions.[10]

The crucial point here is that in all these cases it is always explicitly assumed that the Euler–Lagrange equations can be put in a normal form or, in other words, that the Lagrangians considered are regular; the matrix W of the coefficients of the accelerations is regular.

Finally, the Hamiltonian formulation is introduced by starting with a regular Lagrangian and defining the Legendre transformation, which is invertible. We arrive in this way at the Hamilton equations describing the time evolution of the system. The geometric concept generalizing this approach has been shown to be that of Hamiltonian dynamical systems,[11,12] which are but a triplet (M, Ω, H), where (M, Ω) is a symplectic manifold and H a differentiable function on M. In fact, the equations determining the integral curves of the vector field Γ, defined by $i(\Gamma)\Omega = dH$, are Hamilton-like equations if appropriate coordinates (Darboux coordinates) are chosen.

On the other hand, given the configuration space of a system, it does not necessarily admit global coordinates. Then the Hamilton principle as usually stated is not a geometric concept, because it is the integration of one-forms on paths, and not that of functions, which is meaningful. There is, however, a new method of arriving at the Euler–Lagrange equations for autonomous systems without using the Hamilton principle. It is simply the consideration of the Hamiltonian systems (TQ,ω_L,E_L) defined by a regular Lagrangian.[11,12] On the projection on the basis of the integral curves of the dynamical vector field, the Euler–Lagrange equations hold.

This paper aims to analyze some points that arise when a singular Lagrangian is considered. Then we must decide between choosing either the Hamilton principle approach or the presymplectic-system geometric approach[13–16] as the basic principle generalizing the case of regular systems to singular Lagrangians, both approaches being equivalent in the former case. Instead of considering the "dogma" of the Hamilton principle and the corresponding Euler–Lagrange equations, we will follow the alternative approach. We will find that the dynamical equation admits solutions that have nothing to do with the Euler–Lagrange equations, the latter arising only when we deal with special kinds of vector fields, the so-called second-order differential equation fields (hereafter abbreviated SODE).

At first it might seem that the Euler–Lagrange equations have been well supported by the coincidence of its predictions with experimental results for a long time. But it is worth recalling that singular Lagrangians have associated gauge degrees of freedom that are fictitious, so we have no reason for insisting that the Euler–Lagrange equations are the ones describing the evolution of such gauge degrees. On the contrary, reduction of the presymplectic system is possible, and we will find in any case a Hamiltonian dynamical

 0022-2488/88/051143-07$02.50

system describing the reduced system that contains only the true degrees of freedom.

The paper is organized as follows: Sec. II presents notation and basic definitions in a brief introduction to the geometric approach. The SODE problem is analyzed in Sec. III, and the relation of the Euler–Lagrange equations with the geometric formulation is explained. In Sec. IV we study whether the image of ker ω_L under the vertical endomorphism covers ker FL_*. This property will be used in Sec. V to explain how additional SODE constraints will arise with respect to the set of conditions derived by application of the algorithm developed by Gotay *et al.*[13,14] The former constraints will be shown to be non-FL-projectable in Sec. VI, and the connection between Lagrangian and Hamiltonian constraints is also given, completing some results obtained very recently.[17–19] Finally, Sec. VII contains a group of examples that are very useful for illustrating the theory.

II. NOTATION AND BASIC DEFINITIONS

The geometric approach for the Lagrangian description of an autonomous mechanical system makes use of a differentiable manifold Q as the configuration space and its tangent bundle TQ as the velocity-phase space. We recall that such a vector bundle has a canonical $(1,1)$ tensor called the vertical endomorphism,[20] whose coordinate expression is given by

$$S=\frac{\partial}{\partial v^i}\otimes dq^i. \tag{2.1}$$

Given a function $L\in C^\infty(TQ)$, we may define a function $E_L=\Delta(L)-L$, where $\Delta\in\mathscr{X}(TQ)$ denotes the Liouville vector field generating dilatations along the fibers of TQ, and an exact two-form $\omega_L=-d(dL\circ S)$, which in coordinates of TQ are written

$$\theta_L=\frac{\partial L}{\partial v^i}dq^i, \tag{2.2}$$

$$\omega_L=-d\theta_L$$

$$=\frac{\partial^2L}{\partial v^i\partial q^j}dq^i\wedge dq^j-\frac{\partial^2L}{\partial v^i\partial v^j}dv^i\wedge dq^j, \tag{2.3}$$

$$E_L=v^i\frac{\partial L}{\partial v^i}-L. \tag{2.4}$$

If ω_L is of a constant rank, then L is called the Lagrange function, E_L the energy function, θ_L the Euler–Poincaré one-form, and ω_L the Lagrange two-form.

The map $\widehat{\omega}_L:\mathscr{X}(TQ)\to\Lambda^1(T^*Q)$, defined by contraction, i.e.,

$$\widehat{\omega}_L(X)Y=\omega_L(X,Y),\quad \forall X,Y\in\mathscr{X}(TQ),$$

is represented by the matrix

$$\widehat{\omega}_L=\begin{bmatrix}A & -W\\ W & 0\end{bmatrix}, \tag{2.5}$$

where we have used a matrix notation, the elements of the matrices A and W being

$$A_{ij}=\frac{\partial^2L}{\partial q^i\partial v^j}-\frac{\partial^2L}{\partial v^i\partial q^j} \tag{2.6}$$

and

$$W_{ij}=\frac{\partial^2L}{\partial v^i\partial v^j}, \tag{2.7}$$

respectively. The particular case of L, where ω_L is of maximal rank, i.e., $\widehat{\omega}_L$ is invertible, is that of regular Lagrangians, and the Legendre transformation[11] is then a local diffeomorphism. Taking into account that det $\widehat{\omega}_L=(\det W)^2$, we see that this situation is characterized by det $W\neq0$.

We recall the geometric meaning of the Legendre transformation. It is a map, FL: $TQ\to T^*Q$, defined by the Lagrangian L, which in coordinates reads

$$p_i\circ FL(v)=\frac{\partial L}{\partial v^i}, \tag{2.8}$$

where the natural coordinates of the cotangent bundle are denoted (q^i,p_i). It is worthy of mention that the differentiable map FL pulls back the canonical one-form $\theta_0=p_i\,dq^i$ onto the Euler–Poincaré one-form θ_L, and therefore the canonical symplectic form Ω in T^*Q onto the Legendre two-form ω_L. Consequently, FL is a local diffeomorphism if and only if ω_L is symplectic. When FL is a global diffeomorphism the Lagrangian is said to be hyper-regular, and an equivalent Hamiltonian formulation is obtained by defining the Hamiltonian H by $H=E_L\circ FL^{-1}$.

III. THE SODE PROBLEM

If the Lagrangian L is regular, it is easy to see that the Lagrangian vector field Γ_L, uniquely defined by the dynamical equation

$$i(\Gamma)\omega_L=dE_L, \tag{3.1}$$

is a second-order differential equation,[20] $S(\Gamma_L)=\Delta$. That is, Γ_L is written in coordinates as

$$\Gamma_L=v^i\frac{\partial}{\partial q^i}+b^i\frac{\partial}{\partial v^i}, \tag{3.2}$$

where the b^i satisfy

$$A\mathbf{v}-W\mathbf{b}=\nabla_{\mathbf{q}}E_L. \tag{3.3}$$

This is because the equations determining the integral curves of Γ_L are

$$\begin{bmatrix}A & -W\\ W & 0\end{bmatrix}\begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix}=\begin{bmatrix}\nabla_{\mathbf{q}}E_L\\ W\mathbf{v}\end{bmatrix} \tag{3.4}$$

which split into two subsystems, (3.3) and

$$W(\mathbf{a}-\mathbf{v})=\mathbf{0}. \tag{3.5}$$

Now the regularity of W means that $\mathbf{a}=\mathbf{v}$ is the only solution of (3.5), and if M denotes the inverse matrix of W, the values of $\mathbf{b}$ are determined by

$$b^i=M^{ij}\alpha_j\frac{\partial}{\partial v^i}, \tag{3.6}$$

where α_j denotes

$$\alpha_j=\frac{\partial L}{\partial q^j}-v^k\frac{\partial^2L}{\partial v^j\partial q^k}. \tag{3.7}$$

The integral curves of Γ_L will be determined by

$$\dot q^i=v^i,\quad \dot v^i=M^{ij}\alpha_j, \tag{3.8}$$

and therefore on the projection on Q of these integral curves the Euler–Lagrange equations hold. The point to be made now is that when L is singular, the system (3.4) may have no solution on some points. Gotay *et al.*[13,14] have developed a geometric algorithm for the determination of a maximal submanifold, called the final constraint submanifold C, in which the dynamical equation (3.1) has a consistent solution. The algorithm generates a decreasing sequence $\{P_k\}$ of submanifolds (with $P_0 = TQ$), and then C is the limit of such a sequence (provided it exists). The restricted equation

$$(i(\Gamma)\omega_L - dE_L)_{|C} = 0 \tag{3.9}$$

has solutions tangent to C, but they are not SODE in the general case. The conditions for the existence of a solution such that it is the restriction of a SODE will lead to a smaller submanifold. In fact, in those points in which the dynamics has a solution, the dynamics presents an ambiguity, where the general solution is $\Gamma_0 + \ker \omega_L$, with Γ_0 a particular solution. The coordinate expression is

$$\Gamma = (v^i + \xi^i)\frac{\partial}{\partial q^i} + b^i \frac{\partial}{\partial v^i}, \tag{3.10}$$

with ξ such that $W\xi = 0$. Its integral curves will be determined by the system

$$\dot{\mathbf{q}} = \mathbf{v} + \xi, \quad \dot{\mathbf{v}} = \mathbf{b}, \tag{3.11}$$

and therefore the Euler–Lagrange equations are no longer true but become

$$\begin{aligned}\frac{\partial^2 L}{\partial v^i \partial v^j}\ddot{q}^j + \frac{\partial^2 L}{\partial v^i \partial q^k}\dot{q}^k - \frac{\partial L}{\partial q^i} \\ = W_{ij}\xi^j + \frac{\partial^2 L}{\partial q^i \partial v^j}\xi^j.\end{aligned} \tag{3.12}$$

IV. THE SETS ker ω_L AND ker FL_*

We are now interested in the relation between

$$\ker \omega_L = \{Z \in \mathscr{X}(TQ) \mid i(Z)\omega_L = 0\} \tag{4.1}$$

and $\ker FL_*$. As a consequence of the relation $\omega_L = FL^*\Omega$, it is obvious that $\ker FL_* \subset \ker \omega_L$. But if we recall the explicit form of the matrix representing FL_*,

$$FL_{*v} = \begin{bmatrix} I & 0 \\ B & W \end{bmatrix}, \tag{4.2}$$

with $B_{ij} = \partial^2 L/\partial v^i \partial q^j$, we see that

$$\ker FL_* = \ker \omega_L \cap \mathscr{X}^v(TQ) = V(\ker \omega_L),$$

where $\mathscr{X}^v(TQ)$ denotes the subset of the vertical vector fields, i.e, in coordinates, $Z \in \mathscr{X}^v(TQ)$ if and only if $Z = b^i \partial/\partial v^i$ with $b^i \in C^\infty(TQ)$.

The image of $\ker \omega_L$ under the vertical endomorphism S is in $\ker FL_*$, because

$$X = \xi^i \frac{\partial}{\partial q^i} + \eta^j \frac{\partial}{\partial v^j} \in \ker \omega_L$$

if and only if $A\xi = W\eta$ and $W\xi = 0$, and since $S(X) = \xi^i \partial/\partial v^i$, we have, taking into account the expression (4.2) for FL_{*v}, that $S(Z)$ is FL-projectable and $FL_* S(Z) = 0$, $\forall Z \in \ker \omega_L$.

The particular instance in which the image of $\ker \omega_L$ covers $V(\ker \omega_L) = \ker FL_*$ has recently been shown to be an important case.[21] It is then possible to choose on the final constraint submanifold a solution of the dynamical equation that is the restriction of a SODE. The corresponding Lagrangians, when they admit a global dynamics, were called type II Lagrangians in a recent paper[22] where their properties were studied.

In order to show the relevance of the property $S(\ker \omega_L) = V(\ker \omega_L)$, we first remark that given a solution Γ of the dynamical equation, another vector field Γ' will also be a solution of (3.1) if and only if the difference $\Gamma - \Gamma'$ lies in $\ker \omega_L$. Similarly, if Γ is a SODE, Γ' is a SODE if and only if the difference $\Gamma' - \Gamma$ is a vertical field. Thus the idea is to modify a given solution Γ of (3.1) by adding an element in $\ker \omega_L$ in order to obtain a SODE solution of (3.1) also. It will be possible if the difference $S(\Gamma) - \Delta$ is the image under S of an element in $\ker \omega_L$. In fact, we know that Γ must be of the form

$$\Gamma = (v^i + \xi^i)\frac{\partial}{\partial q^i} + b^i \frac{\partial}{\partial v^i},$$

with $W\xi = 0$, and therefore $S(\Gamma) - \Delta = \xi^i \partial/\partial v^i$ lies in $\ker FL_*$. The point is that if such a difference is in the image $S(X)$ of an $X \in \ker \omega_L$, then $\Gamma - X$ is a SODE solution of the dynamics as well.

We recall that if $\mathscr{H}$ is a Hilbert space and T a bounded operator in such space, the closure of the image of T coincides with the orthogonal of the kernel of its adjoint operator $T^\dagger$ (see, for example, Ref. 23, p. 357 or Ref. 24, p. 214). If we consider the particular case where $\mathscr{H}$ is a finite-dimensional Euclidean space and T a (skew-)symmetric operator, we can conclude that if T is a (skew-)symmetric matrix, then the linear system $Tx = y$ has a solution if and only if $\langle z,y \rangle = 0$, $\forall z \in \ker T$, where $\langle \, , \, \rangle$ denotes the Euclidean inner product. Such a solution is not uniquely determined except up to addition of an element of $\ker T$. This fact may be used to prove the following result.

Theorem 1: Let $X = \xi^i \partial/\partial v^i$ be a vector field in $\ker FL_* = V(\ker \omega_L)$. Then there exists $Z \in \ker \omega_L$ such that $S(Z) = X$ if and only if $\langle \xi', A\xi \rangle = 0$, for every ξ' such that $W\xi' = 0$.

Proof: The condition for X to be in $\ker FL_*$ is $W\xi = 0$. There will exist a $Z \in \ker \omega_L$ such that $S(Z) = X$ if and only if the system

$$W\eta = A\xi$$

has a solution. Then the remark preceding the statement of Theorem 1 shows that it is equivalent to $\langle \xi', A\xi \rangle = 0$, $\forall \xi'$ such that $W\xi' = 0$.

Corollary 1: The map S', restriction of S to $\ker \omega_L$, is onto $\ker FL_*$ if and only if $\langle \xi', A\xi \rangle = 0$, $\forall \xi, \xi'$ such that $W\xi' = W\xi = 0$.

V. THE LAGRANGIAN CONSTRAINTS

In this section we will analyze the compatibility conditions for the existence of solutions of the dynamical equations. The system to be considered is

$$\begin{aligned} A\mathbf{a} - W\mathbf{b} &= \nabla_{\mathbf{q}} E_L, \\ W(\mathbf{a} - \mathbf{v}) &= 0, \end{aligned} \tag{5.1}$$

which is the local expression of (3.1) with $\Gamma = a^i \partial/\partial q^i + b^j \partial/\partial v^j$. The general solution of the second system is

$\mathbf{a} = \mathbf{v} + \xi$, with ξ such that $W\xi = \mathbf{0}$, and therefore the first subsystem becomes

$$W\mathbf{b} = \alpha + A\xi, \tag{5.2}$$

where

$$\alpha_i = \frac{\partial L}{\partial q^i} - \frac{\partial^2 L}{\partial v^i \partial q^j} v^j. \tag{5.3}$$

We first analyze the existence of a SODE solution of (3.1), i.e., a solution of (5.2) with $\xi = \mathbf{0}$. We know that the subsystem

$$W\mathbf{b} = \alpha \tag{5.4}$$

has a solution if and only if

$$\langle \xi, \alpha \rangle = 0, \quad \forall \xi \text{ such that } W\xi = \mathbf{0}. \tag{5.5}$$

Then (5.5) are the constraint functions selecting the submanifold S of TQ where Eq. (3.1) has a SODE solution (but it may not be tangent to S).

Had we just looked for a solution of (3.1), not a SODE but a general vector field, we would have had to search for a ξ such that $W\xi = \mathbf{0}$ in such a way that there exists a solution of (5.2). Let us choose a basis $[\xi_\mu]$ of the kernel of W. Then any $\xi \in \ker W$ can be written as a linear combination $\xi = \lambda_\mu \xi_\mu$, with $m = 1,\ldots,R = \dim \ker W$. The conditions for the existence of a solution are now

$$\langle \xi_\mu, W\mathbf{b} \rangle = 0 = \langle \xi_\mu, \alpha \rangle + \lambda_\nu \langle \xi_\mu, A\xi_\nu \rangle, \quad \forall \mu = 1,\ldots,R. \tag{5.6}$$

In the particular case in which $\langle \xi_\mu, A\xi_\nu \rangle = 0$ for any pair of indices, we will again find (5.5) as the constraint functions (ξ remains absolutely undetermined in $\ker W$). But when the rank r of the $R \times R$ matrix $\langle \xi_\mu, A\xi_\nu \rangle$ is greater than zero, there will be $R - r$ linearly independent combinations

$$b_{\tau\mu} \langle \xi_\mu, A\xi_\nu \rangle = 0. \tag{5.7}$$

From these we will find the constraint functions

$$b_{\tau\mu} \langle \xi_\mu, \alpha \rangle = 0, \quad \forall \tau = 1,\ldots,R - r, \tag{5.8}$$

while r of the values of the parameters λ are determined by (5.6) in terms of the remaining $R - r$ values. The freedom in the choice of the basis of $\ker \omega_L$ allows us to redefine a new basis, in which the $R - r$ first elements are the combinations

$$\gamma_\tau = b_{\tau\mu} \xi_\mu, \quad \tau = 1,\ldots,R - r, \tag{5.9}$$

in such a way that the dynamical constraints are just the $R - r$ functions

$$\langle \gamma_\tau, \alpha \rangle = 0, \quad \forall \tau = 1,\ldots,R - r. \tag{5.10}$$

The adjective "dynamical" indicates that constraints have nothing to do with the SODE condition but only with the existence of a solution for (3.1).

It is noteworthy that the new basis is such that $\langle \gamma_\tau, A\xi_\mu \rangle = 0$, for any index μ, and therefore the vector field $\gamma^i_\tau \, \partial/\partial v^i$ is such that there exists a $Z \in \ker \omega_L$ with $S(Z) = \gamma^i_\tau \, \partial/\partial v^i$. So we have essentially proved the following result.

Theorem 2: Let L be a singular Lagrangian. Then there exists a basis $\{\gamma_\tau, \xi_\mu\}$ (with $\tau = 1,\ldots,R - r$ and $\mu = R - r + 1,\ldots,R$) of $\ker W$ such that (i) the dynamical Lagrangian constraints are given by

$$\langle \gamma_\tau, \alpha \rangle = 0, \quad \forall \tau = 1,\ldots,R - r; \tag{5.11}$$

and (ii) the constraints for the existence of a SODE solution of the dynamics are (5.11) together with

$$\langle \xi_\mu, \alpha \rangle = 0, \quad \forall \mu = R - r + 1,\ldots,R. \tag{5.12}$$

Note, however, that in some cases the constraints may be reduced to identities.

In order to understand better the meaning of the constraint functions, we remark that if $Z \in \ker \omega_L$ is written as

$$Z = \xi^i \frac{\partial}{\partial q^i} + \eta^j \frac{\partial}{\partial v^j},$$

then $\langle \xi, \alpha \rangle = -Z(E_L)$. In fact,

$$\begin{aligned} ZE_L &= \xi^i \frac{\partial}{\partial q^i}\left(v^j \frac{\partial L}{\partial v^j} - L\right) + \eta^k \frac{\partial}{\partial v^k}\left(v^j \frac{\partial L}{\partial v^j} - L\right) \\ &= \xi^i \left[\frac{\partial^2 L}{\partial q^i \partial v^j} v^j - \frac{\partial L}{\partial q^i}\right] + \eta^k \frac{\partial^2 L}{\partial v^j \partial v^k} v^j, \end{aligned} \tag{5.13}$$

and the condition $Z \in \ker \omega_L$ means

$$\frac{\partial^2 L}{\partial v^j \partial v^k} \eta^k = \left[\frac{\partial^2 L}{\partial q^j \partial v^k} - \frac{\partial^2 L}{\partial q^k \partial v^j}\right] \xi^k,$$

which when substituted into (5.13) leads to

$$ZE_L = \xi^k \left[\frac{\partial^2 L}{\partial v^k \partial q^j} v^j - \frac{\partial L}{\partial q^k}\right] = -\langle \xi, \alpha \rangle.$$

Consequently, the conditions (5.11) are simply

$$ZE_L = 0, \quad \forall Z \in \ker \omega_L. \tag{5.14}$$

The above expression shows that the conditions obtained for vertical vector fields $Z \in V(\ker \omega_L)$ reduce to identities.[25] Moreover, we can give a new expression that holds not only for dynamical constraints but also for SODE conditions. It is

$$\langle i(\Gamma_0)\omega_L - dE_L, Z \rangle = 0, \tag{5.15}$$

with Z any vector field such that $S(Z) \in V(\ker \omega_L)$ and Γ_0 an arbitrary SODE. This expression becomes (5.14) when Z is chosen in $\ker \omega_L$ as assumed in the derivation of (5.14).

VI. THE CONNECTION WITH THE HAMILTONIAN FORMULATION

When L is a singular Lagrangian, the Legendre transformation $FL: TQ \to T^*Q$ is not a local diffeomorphism. We will only consider here the case in which L is an "almost regular" Lagrangian, according to the terminology used by Gotay and Nester[14]; FL is a submersion onto its image and the fibers $FL^{-1}\{FL(v)\}$ are assumed to be connected. Then $\ker FL_*$ is an involutive distribution that generates a foliation $\mathscr{F}$ in TQ, and the quotient space $TQ/\mathscr{F}$ is a differentiable manifold that is canonically equivalent to the submanifold M_0, the primary constraint submanifold in Dirac's terminology.[26,27] If $C^0(T^*Q, M_0)$ denotes the set of constraint functions for M_0, then

$$FL^*(C^0(T^*Q, M_0)) = C, \tag{6.1}$$

where C denotes the set of constant functions on TQ.

The main point is the existence of a correspondence $R(L): \mathscr{X}(T^*Q) \to \mathscr{X}^v(TQ)$, which is defined in a pointwise sense by[28]

$$[R(L)X](v) = \xi^v[\tau_{*FL(v)}X], \quad \forall X \in \mathscr{X}(T^*Q), \tag{6.2}$$

where τ: $T^*Q \to Q$ is the projection of the cotangent bundle and ξ^v the vertical lift ξ^v: $T_{\pi(v)}Q \to T_v(TQ)$, given by[20]

$$\xi^v(w)f = \frac{d}{dt}\{f(v+tw)\}_{|t=0}, \quad \forall f \in C^\infty(v).$$

The expression in coordinates of $R(L)X$ with $X = a^i\partial/\partial q^i + b^i\partial/\partial v^i$ is

$$[R(L)X](v) = a^i(FL(v))\left(\frac{\partial}{\partial v^i}\right)_{|v}. \tag{6.3}$$

In particular, if $f \in C^\infty(T^*Q)$ and the vector field $X_f \in C^\infty(T^*Q)$ is defined by

$$i(X_f)\Omega = df, \tag{6.4}$$

then the vector field $R(L)X_f$ is given by

$$[R(L)X_f](v) = \left(\frac{\partial f}{\partial p_i}\right)_{Fl(v)}\left(\frac{\partial}{\partial v^i}\right)_{|v}. \tag{6.5}$$

The point to be stressed here is that when ϕ is a constraint function for M_0, $R(L)X_\phi$ lies in ker FL_*. In fact, this is a straightforward consequence of (6.1) because, for any $Y \in \mathscr{X}(TQ)$, $Y(\phi \circ FL) = 0$, and when Y is taken to be one of the vertical fields, with $Y = \partial/\partial v^i$, it becomes

$$0 = \frac{\partial}{\partial v^i}[\phi \circ FL] = FL_{*v}\left(\frac{\partial}{\partial v^i}\right)\phi = W_{ij}(v)\left(\frac{\partial \phi}{\partial p_j}\right)_{|FL(v)}. \tag{6.6}$$

Then (6.5) with this expression shows that $R(L)X_\phi \in \ker FL_*$. Moreover, not only is

$$\{R(L)X_\phi | \phi \in C^0(T^*Q, M_0)\} \subset \ker FL_*,$$

but both sides coincide, too, as a simple counting of dimensions shows.

Theorem 3: If ϕ and ϕ' are two constraint functions $\phi, \phi' \in C^0(T^*Q, M_0)$, then

$$\{\phi,\phi'\}_{FL(v)} = -\langle \xi_\phi, A\xi_{\phi'}\rangle_v = \omega_L(Y_\phi, Y_{\phi'})_v, \tag{6.7}$$

with Y_ϕ a vector field such that

$$S(Y_\phi) = R(L)X_\phi = \xi^i_\phi\,\partial/\partial v^i,$$

and similarly for ϕ'.

Proof: This is just a matter of checking, because

$$\langle \xi_\phi, A\xi_{\phi'}\rangle_v = \left(\frac{\partial \phi}{\partial p_j}\right)_{FL(v)} \times \left\{\frac{\partial^2 L}{\partial q^j\,\partial v^k} - \frac{\partial^2 L}{\partial v^j\,\partial q^k}\right\}_v \left(\frac{\partial \phi}{\partial p_k}\right)_{FL(v)},$$

and therefore

$$\langle \xi_\phi, A\xi_{\phi'}\rangle = \left(\frac{\partial \phi}{\partial p_j}\right)_{FL(v)}\left(\frac{\partial \phi'}{\partial q_j}\right)_{FL(v)} - \left(\frac{\partial \phi}{\partial q_j}\right)_{FL(v)}\left(\frac{\partial \phi'}{\partial p_j}\right)_{FL(v)}.$$

On the other hand, no matter what the choice of arbitrary η and η', if Y_ϕ is given by $Y_\phi = \xi^i_\phi\,\partial/\partial q^i + \eta^i_\phi\,\partial/\partial v^i$ and similarly for $Y_{\phi'}$, we can check that $\omega_L(Y_\phi, Y_{\phi'})_v = -\langle \xi_\phi A\xi_{\phi'}\rangle_v$, because of $W\xi_\phi = W\xi_{\phi'} = 0$.

For any constraint function ϕ, let Y_ϕ be, as in Theorem 3, a vector field $Y_\phi \in \mathscr{X}(TQ)$ such that $S(Y_\phi) = R(L)X_\phi$, with X_ϕ as in (6.4). Then the function $i(Y_\phi)(i(\Gamma_0)\omega_L - dE_L)$, whose coordinate expression is $\langle \xi_\phi, \alpha\rangle$, is a constraint function in TQ. Moreover, the vector field Y_ϕ can be chosen in ker ω_L if and only if ϕ is of the first class (at the M_0 level), or in other words, X_ϕ is tangent to M_0, as a consequence of Theorems 2 and 3. These last constraint functions are just the dynamical constraints, while the remaining constraint functions, the second-class (at the M_0 level) primary constraint functions, will be associated to SODE conditions.

Other results that may be straightforwardly deduced from the relation (6.7) in Theorem 3 are summed up in the following theorem.[22]

Theorem 4: (i) All the primary constraint functions in T^*Q are of the first class (M_0 called coisotropic) if and only if there are no SODE conditions, i.e., all the primary constraints in the Lagrangian formulation are dynamical ones.

(ii) If all the primary constraints are of the second class (M_0 said to be symplectic), then all the primary constraint functions in the Lagrangian formulation are SODE conditions, and there are no dynamical constraints.

We also remark that in this last situation there will not exist secondary constraints in the phase-space formulation. In fact, the comment after formula (5.14) shows that the energy function is FL-projectable, that is, there exists a function $H \in C^\infty(M_0)$ such that $H \circ FL = E_L$. Let the submanifold j: $M_0 \to T^*Q$ be symplectic. If X is the vector field solution of the equation $i(X)j^*\Omega = dH$, then j_*X is a solution of $j^*\{i(Y)\Omega\} = dH$ that is just tangent to M_0, and consequently there are no secondary constraints in the Hamiltonian formulation.

Besides the aforementioned relationship between the dynamical or SODE Lagrangian constraints and the first- or second-class primary constraints, respectively, there exists a well-known correspondence between the dynamical Lagrangian constraints and the secondary Hamiltonian ones, obtained by making use of the pullback FL^*. Therefore there is a local basis[14] of FL-projectable Lagrangian constraints defining the submanifold P_1 of TQ such that $FL(P_1) = M_1$. On the contrary, if S_1 is the submanifold defined by both the dynamical and the SODE Lagrangian constraints, it can be seen that $FL(S_1) = FL(P_1) = M_1$. (The proof is similar to that of Proposition 3 in the paper by Gotay and Nester.[15]) It means that every SODE Lagrangian constraint is not FL-projectable. Finally, while ker FL_* is tangent to P_1, we can only assert the S_1 tangency for the elements of $S(\ker \omega_L)$. Moreover, if there is no SODE Lagrangian constraint trivial on P_1, then each $Z \in \ker FL_* - S(\ker \omega_L)$ will be not tangent to S_1.

VII. EXAMPLES

In this section we shall carefully examine, using the methods developed in this paper, some simple examples that were discussed in several previous papers. In the first place, we should mention an example in which there are no SODE conditions. This will be the case when $\langle \xi', A\xi\rangle = 0$, for any pair of vectors in ker W—for instance, if it is one-dimensional.

Example 1: The Lagrangian L is defined in $T\mathbb{R}^3$ as follows[29]:

$$L(\mathbf{x},\mathbf{v}) = v_1 v_3 + \tfrac{1}{2}(x_2{}^2 x_3).$$

The matrix W is given by

$$W = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

and so ker W is one-dimensional. The only primary constraint is a dynamical one, $\alpha_2 = x_2 x_3 = 0$. The general form of the dynamical vector field on this manifold is

$$\Gamma = v_1 \frac{\partial}{\partial x_1} + v_3 \frac{\partial}{\partial x_3} + \frac{1}{2}\left(x_2{}^2 \frac{\partial}{\partial v_1}\right) + \lambda \frac{\partial}{\partial x_2} + \mu \frac{\partial}{\partial v_2},$$

because ker ω_L is generated by $\partial/\partial x_2$ and $\partial/\partial v_2$. The SODE condition just fixes the value of the parameter λ as v_2. The energy function is $E_L = v_1 v_3 - \tfrac{1}{2}(x_2{}^2 x_3)$. In the corresponding Hamiltonian formulation there will arise a primary constraint, $\varphi_1 = p_2$, and a secondary one, $\varphi = x_2 x_3$.

A similar case is the one studied by Frenkel,[30] but a word of caution is needed because L is not a true Lagrangian function. The rank of ω_L is not constant and reduces to zero on the submanifold $v_3 = 0$.

The examples of DiStefano,[31]

$$L = \tfrac{1}{2}(v_x{}^2) + \tfrac{1}{2}(x^2 y),$$

Schafir,[32]

$$L = \tfrac{1}{2}(v_x - x^2 + y v_z{}^2),$$

and Cawley,[29]

$$L = v_x v_y + \tfrac{1}{2}(y v_z{}^2),$$

look very similar to the preceding case.

Another similar example is that proposed by Christ and Lee,[33] which in a slightly modified form has recently been analyzed by Nardelly and Soldati[34]:

$$L = \tfrac{1}{2}[\dot{r}^2 + r^2(\dot{\theta} - z)^2] - V(r) \quad (r \neq 0),$$

because ker W is also one-dimensional, and then the primary constraint arising in this case,

$$\alpha_3 = r^2(\dot{\theta} - z) = 0,$$

is dynamical. The kernel of ω_L is now generated by the vector fields $\partial/\partial z + \partial/\partial\dot{\theta}$ and $\partial/\partial\dot{z}$, and there are no secondary constraints in the Lagrangian formulation. In fact, the general solution of the dynamics is

$$\Gamma_{\alpha,\beta} = \dot{r}\frac{\partial}{\partial r} + \dot{\theta}\frac{\partial}{\partial \theta} - (r\dot{\theta}^2 + V'(r) - rz^2)\frac{\partial}{\partial \dot{r}} + \alpha\left[\frac{\partial}{\partial z} + \frac{\partial}{\partial \dot{\theta}}\right] + \beta\frac{\partial}{\partial \dot{z}},$$

and $\Gamma_{\alpha,\beta}(\dot{\theta} - z) = 0$. Thus there are no secondary constraints in this approach. In the Hamiltonian counterpart in which the momenta are given by

$$p_r = \dot{r}, \quad p_\theta = r^2(\dot{\theta} - z), \quad p_z = 0,$$

we will find a primary constraint $p_z = 0$ and a secondary constraint $p_\theta = 0$, but both of the first class. This last fact corresponds to the nonexistence of SODE constraints.

Finally, we note that there exist cases with no SODE restrictions in which the dimension of ker W is greater than 1. In the following example, one proposed by Sundermeyer[35] but slightly modified in order to remove the inconsistency, the restriction of A onto ker W vanishes:

$$L = \tfrac{1}{2}(q_1 v_2{}^2) + q_2 q_3 \quad (q_1 \neq 0).$$

There are only dynamical constraints,

$$\alpha_1 = v_2 = 0 \quad \text{and} \quad \alpha_3 = q_2 q_1 = 0.$$

Example 2: Let us consider now a case in which there are no dynamical constraints except SODE conditions. An example is the one given by Schafir,[32]

$$L = \tfrac{1}{2}(v_1{}^2 - x_1{}^2) + x_2 v_3.$$

The matrix W is now

$$W = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

while the matrix $\langle \xi_\mu, A\xi_\nu \rangle$ is regular. Therefore there are no dynamical constraints except the following SODE restrictions:

$$\alpha_2 = v_3 = 0 \quad \text{and} \quad \alpha_3 = v_2 = 0,$$

the kernel of ω_L being generated by the vector fields $\partial/\partial v_2$ and $\partial/\partial v_3$. The general solution of the dynamics is

$$\Gamma = v_1 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial v_1} + \lambda \frac{\partial}{\partial v_2} + \mu \frac{\partial}{\partial v_3}.$$

The Hamiltonian counterpart has two primary constraints,

$$\varphi_1 = p_2 \quad \text{and} \quad \varphi_2 = p_3 - q_2,$$

which are of the second class at this level. This situation corresponds to the case studied in (ii) of Theorem 4, and, as indicated there, neither dynamical primary constraints in the Lagrangian formulation nor secondary ones in the Hamiltonian approach will arise. The same assertion is true for the analogous case studied by Nesterenko and Chervyakov,[36]

$$L = \tfrac{1}{2}(v_1{}^2) - v_2 x_3.$$

Example 3: A simple example of mixed type with both dynamical and SODE restrictions is

$$L = v_1{}^2 + q_3 v_4 + q_2{}^2 + q_4{}^2,$$

in which ker W is three-dimensional. There is a dynamical constraint,

$$\alpha_2 = q_2 = 0,$$

but there are two SODE restrictions as well,

$$\alpha_3 = v_4 = 0 \quad \text{and} \quad \alpha_4 = q_4 = 0.$$

In the Hamiltonian approach, three primary constraints corresponding to the three dimensions of ker W,

$$\phi_1 = p_2 = 0, \quad \phi_2 = p_3 = 0, \quad \phi_3 = p_4 - q_3 = 0,$$

and just one secondary constraint,

$$\phi_4 = q_2 = 0,$$

will arise, since there is one dynamical constraint in the Lagrangian formulation. The two SODE restrictions have no counterpart in the Hamiltonian formulation.

ACKNOWLEDGMENTS

One of the authors (C. L.) thanks Diputación General de Aragón for a grant. Partial financial support by CICYT is also acknowledged.

[1]E. J. Saletan and A. H. Cromer, *Theoretical Mechanics* (Wiley, New York, 1971).
[2]V. I. Arnold, *Les méthodes mathématiques de la Méchanique Classique* (Mir, Moscow, 1976).
[3]E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974).
[4]I. M. Anderson, *Aspects of the Inverse Problem to the Calculus of Variations*, Proceedings of the Conference on Differential Geometry and Its Applications, Brno, 1986 (Reidel, Dordrecht, 1987).
[5]W. Sarlet, J. Phys. A: Math. Gen. **15**, 1503 (1982).
[6]W. Sarlet, *Proceedings of the IUTAM–SIMM Symposium on Modern Developments in Analytical Mechanics* (Accademia delle Science, Torino, 1983), pp. 737–751.
[7]W. Sarlet and F. Cantrijn, SIAM Rev. **23**, 467 (1981).
[8]M. Henneaux, Ann. Phys. (NY) **140**, 45 (1982).
[9]M. Crampin, J. Phys. A: Math. Gen. **14**, 2567 (1981).
[10]P. Havas, Nuovo Cimento Suppl. **5**, Ser. X, 363 (1957).
[11]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Reading, MA, 1978), 2nd ed.
[12]G. Marmo, E. J. Saletan, A. Simoni, and B. Vitale, *Dynamical Systems, A Differential Geometric Approach to Symmetry and Reduction* (Wiley, Chichester, 1985).
[13]M. J. Gotay, J. M. Nester, and G. Hinds, J. Math. Phys. **19**, 2388 (1978).
[14]M. J. Gotay and J. M. Nester, Ann. Inst. H. Poincaré A **30**, 129 (1979).
[15]M. J. Gotay and J. M. Nester, Ann. Inst. H. Poincaré A **32**, 1 (1980).
[16]J. F. Cariñena, J. Gomis, L. A. Ibort, and N. Román, J. Math. Phys. **26**, 1961 (1985).
[17]C. Batlle, J. Gomis, J. M. Pons, and N. Román, J. Math. Phys. **27**, 2953 (1986).
[18]C. Batlle, J. Gomis, J. M. Pons, and N. Román, Lett. Math. Phys. **13**, 17 (1987).
[19]C. Batlle, J. Gomis, J. M. Pons, and N. Román, J. Phys. A: Math. Gen. **20**, 5113 (1987).
[20]M. Crampin, J. Phys. A: Math. Gen. **16**, 3755 (1983).
[21]J. F. Cariñena and L. A. Ibort, J. Phys. A: Math. Gen. **18**, 3335 (1985).
[22]F. Cantrijn, J. F. Cariñena, M. Crampin, and L. A. Ibort, J. Geom. and Phys. **3**, 353 (1986).
[23]A. W. Naylor and G. R. Sell, *Linear Operator Theory* (Holt, Rinehart and Winston, New York, 1971).
[24]R. D. Milne, *Applied Functional Analysis* (Pitman, Boston, 1980).
[25]F. Cantrijn, *On the Geometry of Degenerate Lagrangians*, Proceedings of the Conference on Differential Geometry and its Applications, Brno, 1986 (Univ. J. E. Parkyne, Brno, 1987).
[26]P. A. M. Dirac, Can. J. Math. **2**, 129 (1950).
[27]P. A. M. Dirac, *Lectures on Quantum Mechanics*, Belfer Graduate School of Science Monograph Series, No. 2 (Yeshiva University, New York, 1964).
[28]J. F. Cariñena, C. López, and N. Román, "Geometric study of the connection between the Lagrangian and Hamiltonian constraints," J. Geom. Phys. (in press).
[29]R. Cawley, Phys. Rev. Lett. **42**, 413 (1979).
[30]A. Frenkel, Phys. Rev. D **21**, 2986 (1980).
[31]R. Di Stefano, Phys. Rev. D **27**, 1752 (1983).
[32]R. L. Schafir, J. Phys. A: Math. Gen. **15**, L331 (1982).
[33]N. H. Christ and T. D. Lee, Phys. Rev. D **22**, 939 (1980).
[34]G. Nardelli and R. Soldati, Nuovo Cimento A **91**, 319 (1986).
[35]K. Sundermeyer, "Constrained dynamics," in *Lecture Notes in Physics*, No. 169 (Springer, Berlin, 1982).
[36]V. V. Nesterenko and A. M. Chervyakov, Theor. Math. Phys. **64**, 701 (1985).

Symplectic geometry and integrable m-body problems on the line

J. P. Francoise
Bat. de Mathématiques n° 425, Université de Paris-sud, Orsay, 91405 Cedex, France

(Received 19 November 1987; accepted for publication 20 January 1988)

A conjecture of Gallavotti and Marchioro [J. Math. Anal. Appl. **44**, 661 (1973)] is proved by using symplectic techniques.

I. INTRODUCTION

The Hamiltonian of the classical system that we consider here is

$$H=\frac{1}{2}\sum_{i=1}^{m} y_i^2+\frac{g^2}{2}\sum_{i\neq j}\frac{1}{(x_i-x_j)^2}+\frac{1}{2}\lambda^2\sum_{i=1}^{m} x_i^2,$$

for the symplectic form $\omega=\Sigma_{i=1}^m dx_i\wedge dy_i$ on the manifold T^*W,

$$W=\{X\in\mathbb{R}^m | x_i-x_j\neq 0,\ i\neq j\}.$$

The corresponding quantum operator is

$$\hat{H}=\frac{h^2}{2}\sum_{i=1}^{m}\frac{\partial^2}{\partial x_i^2}+\frac{g^2}{2}\sum_{i\neq j}\frac{1}{(x_i-x_j)^2}+\frac{\lambda^2}{2}\sum_{i=1}^{m}x_i^2.$$

This m-body model on the line has been studied by Calogero[1] and Sutherland,[2] and in the classical case by Adler,[3] Calogero,[1] Kazhdan, Kostant, and Sternberg,[4] Moser,[5] and Olshanetsky and Perelomov.[6]

The classical system is completely integrable in the Liouville sense. The level hypersurfaces of the Hamiltonian are compact and so there are invariant tori; we propose to compute the action-angle coordinates for this model.

In the quantum case the Hamiltonian has a discrete spectrum which has been determined by Calogero.[1] Gallavotti and Marchioro[7] then proved that the semigroup generated by H can be computed by the Feynmann–Kac formula and they deduced the limit $\lim_{h\to 0}(2\pi h)^m\,\mathrm{Tr}(\exp-\beta\hat{H})$. From the expression they found for the integral $\int dx\,dy(\exp-\beta H)$, they proposed a conjecture on the classical system. We use here the Arnol'd formula and the theorem of convexity of the image of the moment map[8,9] of a symplectic action of the torus to prove the conjecture as it was announced in Ref. 10.

Theorem: There is a global symplectic transformation, defined on the complement of an analytic set of codimension 2, Σ,

$$F\colon T^*W-\Sigma\to\mathbb{R}_+^m\times\mathbb{T}^m$$
$$(x,y)\to(\eta,\xi)$$

so that

$$F^*H=\lambda\sum_{i=1}^{m}k\eta_k+\frac{\lambda g m(m-1)}{2},$$

for $g>0$.

II. THE LAX PAIR

Hamilton's equation $\iota_x\omega=dH$ defines a vector field whose flow is a solution of

$$\dot{x}_i=\frac{\partial H}{\partial y_i}=y_i,\qquad \dot{y}_i=-\frac{\partial H}{\partial x_i}=\lambda^2 x_i+2\sum_{j\neq i}\frac{g^2}{(x_i-x_j)^3}.\tag{1}$$

Adler and Moser proved that (1) implies

$$\dot{L}=[A,L]-\lambda^2X\quad\text{and}\quad\dot{X}=[A,X]+L,\tag{2}$$

with

$$L_{ij}=y_i\delta_{ij}+\sqrt{-1}g/(x_i-x_j)(1-\delta_{ij}),\quad X_{ij}=x_i\delta_{ij},$$
$$A_{ij}=\sqrt{-1}\delta_{ij}\sum_{k\neq i}\frac{g}{(x_i-x_k)^2}+\frac{(1-\delta_{ij})(g\sqrt{-1})}{(x_i-x_j)^2}.\tag{3}$$

We define a matrix $U(t)$ (Ref. 5) such that

$$\dot{U}=AU\quad\text{and}\quad U(0)=Id.\tag{4}$$

We use $Y(t)=U(t)XU(t)^{-1}$, the conjugate of X, by the flow of A.

The Lie formula gives

$$\dot{Y}(t)=U(t)(\dot{X}-[A,X])U(t)^{-1}=ULU^{-1}$$

and

$$\ddot{Y}(t)=U(t)(\dot{L}-[A,L])U(t)^{-1}=-\lambda^2UXU^{-1}.$$

So we get

$$\ddot{Y}(t)=-\lambda^2Y(t),\tag{5}$$

which gives by integration

$$Y(t)=Y(0)\cos\lambda t-(\dot{Y}(0)/\lambda)\sin\lambda t.\tag{6}$$

Equation (6) determines completely the flow of (1) because the positions x_i are the eigenvalues of the matrix $Y(t)$.

From (6), we easily deduce that

$$Y(t)^2+(1/\lambda^2)\dot{Y}(t)^2=X(0)^2+L(0)^2/\lambda^2,$$

and that the matrix $P=\lambda^2X^2+L^2$ is isospectral.

In fact we can check that

$$\dot{P}=[A,P].\tag{7}$$

We have a Lax pair for the flow. Let us introduce $(F_1,\dots,F_m)$, the eigenvalues of P.

III. THE ACTION ANGLES OF THE CALOGERO–MOSER SYSTEM WITH AN EXTERNAL QUADRATIC FORCE

We follow the presentation of Kazhdan, Kostant, and Sternberg.[4] The group $G=U(m,\mathbb{C})$ defines a symplectic action on the cotangent bundle of its Lie algebra $T^*g\simeq g\oplus g^*\simeq g\oplus g$ equipped with the canonical symplectic form

 0022-2488/88/051150-04$02.50

$$\Omega = \mathrm{Tr}(dX \wedge dY) = \sum_{i,j} dX_{ij} \wedge dY_{ji}. \qquad (8)$$

This symplectic action has a moment map Ψ: $(X,Y) \to [X,Y]$, where $[X,Y]$ is seen as an element of g^* through the identification $g \simeq g^*$.

Kazdhan, Kostant, and Sternberg build the reduced space $\Psi^{-1}(\mu)/G_\mu$, where μ is defined by the matrix $\mu_{ij} = \sqrt{-1}g(1-\delta_{ij})$. Here G_μ is the isotropy subgroup of μ.

A fundamental result of Ref. 4 is the following lemma.

Lemma (Kazhdan, Kostant, and Sternberg): The element X of g can always be diagonalized by a transformation of G_μ.

This lemma implies that the reduced space $\Psi^{-1}(\mu)/G_\mu$ can be parametrized by the couples $(X,L) \in g \times g$ such that

$$X_{ij} = x_i \delta_{ij} \quad \text{and} \quad [X,L] = \mu. \qquad (9)$$

We find that the nondiagonal terms of L are necessarily of the form $L_{ij} = \sqrt{-1}g/(x_i - x_j)$.

The diagonal terms of L left undetermined by relation (9) will be denoted y_i. The symplectic form is then

$$\omega = \mathrm{tr}(dX \wedge dL) = \sum_{i=1}^{m} dx_i \wedge dy_i, \qquad (10)$$

and we find the data (3) with the Hamiltonian

$$H = \tfrac{1}{2}\mathrm{Tr}(\lambda^2 X^2 + L^2).$$

We begin now the computation of the action angles by introducing "matrix polar coordinates."

By relations (3), we see that $\sqrt{-1}X$ and $\sqrt{-1}L$ are elements of the Lie algebra g. We use the matrices (which are not in g)

$$Z = \lambda X + \sqrt{-1}L \quad \text{and} \quad \overline{Z} = \lambda X - \sqrt{-1}L. \qquad (11)$$

Given a matrix M, we denote by M^* the matrix whose elements are the complex conjugates of the corresponding elements of M. We find that

$$\overline{Z} = {}^t Z^*. \qquad (12)$$

We then show that

$$P = Z\overline{Z} + \sqrt{-1}\lambda\mu. \qquad (13)$$

We introduce the matrix S

$$S = Z\overline{Z} = Z\,{}^t Z^*. \qquad (14)$$

The eigenvalues of S are functions of the eigenvalues of P and so they are constants of the motion. This implies that S is also a Lax matrix for the flow.

Let $q_1,\dots,q_m$ be the eigenvalues of Z (note that they are not necessarily real) and $Q = \mathrm{diag}(q_1,\dots,q_m)$. Let σ_1 be the analytic set of real codimension 2 which is the locus where Z has multiple eigenvalues. There is a matrix V analytic and invertible on $T^*W - \sigma_1$ so that

$$VZV^{-1} = Q.$$

Let us denote $p_1,\dots,p_m$ the diagonal elements of $V\overline{Z}V^{-1}$; we have

$$\omega = \mathrm{tr}(dX \wedge dL) = (1/2\lambda\sqrt{-1})\mathrm{tr}(dZ \wedge d\overline{Z}),$$

$$\mathrm{tr}(dZ \wedge d\overline{Z}) = \mathrm{tr}(dVZV^{-1} \wedge dV\overline{Z}V^{-1})$$

$$= \sum_{i=1}^{m} dq_i \wedge dp_i.$$

Let σ_2 be the analytic set defined by $q_1,\dots,q_m = \det Z = 0$. Let us observe that this set is also of real codimension 2. We introduce on $T^*W - \sigma_2$, the matrix $\log(Q) = \mathrm{diag}(\log q_1,\dots,\log q_m)$, by choosing a logarithm on $\mathbb{C}^*$. We observe that we have, on $T^*W - \Sigma(\Sigma = \sigma_1 u \sigma_2)$,

$$(2\sqrt{-1}\lambda)\omega = \sum_{i=1}^{m} d\log q_i \wedge d(p_i q_i)$$

$$= \mathrm{tr}(d\log Q \wedge dVSV^{-1}),$$

$$(2\sqrt{-1}\lambda)\omega = \mathrm{tr}(dV^{-1}\log QV \wedge dS).$$

Now we use the fact that S ($\sqrt{-1}S$ belongs to the Lie algebra g) can be diagonalized by an element T of the unitary group. Let $\Lambda_1,\dots,\Lambda_m$ be the eigenvalues of S; we can write

$$(2\sqrt{-1}\lambda)\omega = \mathrm{tr}(dT\,V^{-1}\log QVT^{-1} \wedge dT\,ST^{-1}),$$

$$(2\sqrt{-1}\lambda)\omega = \sum_{i=1}^{m} d\Lambda_i \wedge d\beta_i,$$

where $\beta_i = TV^{-1}\log QVT_{ii}^{-1}$ appears as well-defined and has a ramification of logarithmic type when the point (x,y) varies along a loop around σ_2. We use in the following the variables $\xi_i = \beta_i/(2\sqrt{-1}\lambda)$ so that

$$\omega = \sum_{i=1}^{m} d\Lambda_i \wedge d\xi_i.$$

We deduce from (6) that all the orbits of the Calogero–Moser system with a quadratic external potential are periodical of period (not necessarily primitive) $2\pi/\lambda$. The equation of the flow can be written using the variables (Λ_i,ξ_i),

$$\dot{\Lambda}_i = 0, \quad \dot{\xi}_i = \tfrac{1}{2}$$

hence

$$\xi_j = \xi_j(0) + \tfrac{1}{2}t.$$

The flow is periodical of period $2\pi/\lambda$, so we must identify $\xi_j(0)$ and $\xi_j(0) + \pi/\lambda$.

This shows that the variables $\vartheta_j = 2\lambda\xi_j$ are angular variables. We have now proved the following proposition.

Proposition: Let us introduce $\Sigma_j = (\frac{1}{2}\lambda)\Lambda_j$; the variables (Σ_j,ϑ_j) define a system of action-angles coordinates for the completely integrable Hamiltonian that we consider. If we use the one-form $\eta = \sum_{j=1}^{m}(1/2\lambda)\Lambda_j\,d\vartheta_j$ and the classes $\gamma_j(c)$ of the paths $\{\Sigma_j$ constant, ϑ_i $(i \neq j)$ constant, and $0 \leqslant \vartheta_j \leqslant 2\pi\}$ in the cohomology group $H_1(F^{-1}(c),\mathbb{Z})$ we can check the Arnol'd formula

$$\sum_j = \int_{\gamma j(c)} \eta.$$

We have that $H = (F_1 + \cdots + F_m)/2$ and the lemma gives $H = (\Lambda_1 + \cdots + \Lambda_m)/2$ by invariance of the trace. We get then $H = \lambda(\Sigma_1 + \cdots + \Sigma_m)$.

IV. PROOF OF THE GALLAVOTTI–MARCHIORO CONJECTURE

The end of the proof of the conjecture is based on an application of the theorem of the convexity of the moment map,[8,9,11] both in its local and global part.

We used the fact already proved in Ref. 12 that the Calogero–Moser system with a quadratic external potential is collective for the symplectic action of the torus. We proceed separately in each component of the Weyl chamber. Let us consider, for instance, the component defined by $x_1 < x_2 < \cdots < x_m$. The function H is convex, so it has only one critical point in this connected component and it is a minimum. It has been determined by Calogero in Ref. 13 as the point C,

$$y_i = 0, \quad x_i = x_i^0 \quad (i\text{th zero of the Hermite polynomial of degree } m).$$

The value of H at this point is $H^0 = \lambda g m(m-1)/2$.

We begin by performing a translation of the action variables

$$\Sigma_i \to \Sigma_i + \Sigma_i^0,$$

where Σ_i^0 is the value of Σ_i at the point C.

After this translation, the point C becomes the origin of the coordinates 0.

We use then the theorem of Ref. 11 (Sec. 32) which shows the existence in a neighborhood of 0 of a system of symplectic coordinates (x,y) so that

$$H = \sum_{k=1}^{m} \mu_k (x_k^2 + y_k^2).$$

The μ_k are necessarily positive since H has a minimum at 0. In Ref. 13 Calogero determined the eigenvalues of the Hessian of H at the minimum as being $\mu_k = k$.

By a local analysis, we see that the actions given by the proposition are obtained from the local data

$$\eta_k = x_k^2 + y_k^2$$

by a transformation J: $(\eta_1,...,\eta_m) \to (\Sigma_1,...,\Sigma_m)$ which belongs to $\mathrm{Sl}(m,\mathbb{Z})$. We deduce from this fact that the local data have $J^{-1}(\Sigma_1,...,\Sigma_m)$ as a global extension that we still denote by $(\eta_1,...,\eta_m)$. The Arnol'd formula and a change of base in $H_1(F^{-1}(c),\mathbb{Z})$ allows us to make explicit this transformation J. Let $\gamma_i \to \gamma_i'$ be the change of base, so that $\gamma_j = \Sigma_{i=j}^m \gamma_1'$, and $\eta \to \eta'$ the change of one-form defined by $\eta' = \eta - \Sigma_{i=1}^{m-1} g \Sigma_i^0 \, d\vartheta_1$ the actions η_i correspond to (η', γ_i').

The variables $(\eta_1,...,\eta_m)$ can be used instead of the variables $(\Sigma_1,...,\Sigma_m)$. The fundamental theorem of Refs. 8 and 9 tells that the image of the moment map is the convex hull of the critical values if the manifold considered is compact. There is no general statement known about the noncompact case. But the convexity theorem is true for systems for which the stationary phase formula is exact.[14] We have shown in Ref. 12 that the stationary phase formula is exact by an adaptation of the Berline–Vergne proof.

We deduce that the range of $(\eta_1,...,\eta_m)$ is a cone of vertex 0.

The intersection of this cone with a neighborhood U of 0 is equal to $\mathbb{R}_+^m \cap U$ since locally $\eta_k = x_k^2 + y_k^2$. Consequently, the cone is $\mathbb{R}_+^m$ and the variables η_k vary independently from 0 to $+\infty$.

The angular variables ξ_k associated to η_k are *a priori* multivalued analytic functions on the open dense set complement of the algebraic set of codimension two Σ. They vary between 0 and 2π on this open dense set. The fact that they may have singularities on a residual set does not prevent from using them to compute the canonical partition function of H. This ends the proof of the conjecture of Gallavotti and Marchioro.[7] An important consequence of the solution of this problem is that the spectrum of the quantum system computed by Calogero[1] coincides with the approximation given by the geometric quantization.

V. CANONICAL PARTITION FUNCTION OF THE HIGHER-ORDER INTEGRALS OF THE MOTION

The formula of the stationary phase gives an approximation to the sum $Z(\beta) = \int_V \exp(-\beta H)\Omega$, where H is a function defined on a manifold V and Ω is a volume form on V. This approximation is exact when V is symplectic and H is collective for a Hamiltonian action of the torus on V. When H is collective, there is a collection of m functions ($2m = \dim V$): $\Sigma_1, \Sigma_2, ..., \Sigma_m$ so that $H = \Sigma_1 + \cdots + \Sigma_m$ and so that the Poisson brackets $\{\Sigma_i, \Sigma_j\}$ vanish and the Hamiltonian flows of each Σ_i for all $i = 1,...,m$, are periodic and there is a common multiple T to the periods of all the periodic orbits.

Let us consider the higher-order integrals of the motion, $H^{(s)} = \Sigma_1^s + \cdots + \Sigma_m^s$. The Hamiltonians $H^{(s)}$ are no longer collective and nevertheless still integrable and we consider the computation of their corresponding canonical partition functions

$$Z^{(s)}(\beta) = \int_V \exp(-\beta H^{(s)})\Omega$$

as an important problem of symplectic geometry.

As a corollary of the proof of the Gallavotti–Marchioro conjecture, we get the following theorem.

Theorem: For the Calogero–Moser with an external quadratic force, the canonical partition functions of the higher-order integrals of motion are equal to

$$Z^{(s)}(\beta) = (2\pi)^m \int_0^\infty d\eta_1 \cdots \int_0^\infty d\eta_m \times \exp\left(-\beta \sum_{k=1}^{m} (\eta_k + \cdots + \eta_m + \Lambda_k^0)^s\right).$$

This shows the contrasting situation between the simple expression of $Z(\beta)$ obtained by the stationary phase formula and the sums corresponding to the higher-order integrals. It suggests that a detailed study of the special functions which are iterated of error-functions type would be interesting in relation to this problem of symplectic geometry.

[1]F. Calogero, "Solution of the one-dimensional n-body problems with quadratic and/or inversely quadratic pair potentials," J. Math. Phys. **12**, 419 (1973).

[2]B. Sutherland, "Exact results for a quantum many body problem in one dimension. II," Phys. Rev. A **5**, 1372 (1972).

[3]M. Adler, "Some finite dimensional integrable systems and their scattering behavior," Commun. Math. Phys. **55**, 195 (1977).

[4]D. Kazhdan, B. Kostant, and S. Sternberg, "Hamiltonian group actions and dynamical systems of Calogero type," Commun. Pure Appl. Math. **31**, 481 (1978).

[5]J. Moser, "Three integrable Hamiltonian systems connected with isospectral deformations," Adv. Math. **16**, 197 (1975).

[6]M. A. Olshanetski and A. M. Perelomov, "Classical integrable finite-di-

mensional systems related to Lie algebras," Phys. Rep. **71**, 313 (1981).
[7]G. Gallavotti and C. Marchioro, "On the calculation of an integral," J. Math. Anal. Appl. **44**, 661 (1973).
[8]M. F. Ativah, "Convexity and commuting Hamiltonians," Bull. London Math. Soc. **14**, 1 (1982).
[9]V. Guillemin and S. Sternberg, "Convexity of the moment mapping," Inv. Math. **67**, 491 (1982).
[10]J.-P. Francoise, "Preuve d'une conjecture de G. Gallavotti et C. Marchioro," C.R. Acad. Sci. Paris **303**, 791 (1986).
[11]V. Guillemin and S. Sternberg, *Symplectic Techniques in Physics* (Cambridge U. P., Cambridge, 1984).
[12]J.-P. Francoise, "Canonical partition functions and the stationary phase formula," to appear in Commun. Math. Phys.
[13]F. Calogero, "Equilibrium configuration of the one dimensional *n*-body problem with quadratic and inversely quadratic pair potentials," Lett. Nuovo Cimento, **20**, 251 (1977).
[14]M. F. Atiyah and R. Bott, "The moment map and equivariant cohomology," Topology **23**, 1 (1984).

The propagator in the generalized Aharonov–Bohm effect

C. H. Oh and C. P. Soo
Department of Physics, Faculty of Science, National University of Singapore, Kent Ridge, Singapore 0511, Republic of Singapore

C. H. Lai
Center for Theoretical Physics, Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 16 April 1987; accepted for publication 30 December 1987)

The nonrelativistic propagator is derived by formulating the generalized Aharonov–Bohm effect, valid for any gauge group in a general multiply connected manifold, as a gauge artifact in the universal covering space. The loop phase factors and the free homotopy propagators arise naturally. An explicit expression for the propagator when there are two solenoids present is given.

I. INTRODUCTION

The Aharonov–Bohm (AB) effect[1] has long been an intriguing and much discussed topic in the literature. Partly based on this effect Wu and Yang[2] concluded that electromagnetism is a gauge-invariant manifestation of the nonintegrable phase factor; they have furthermore generalized the effect for the non-Abelian gauge fields. A quantitative description of this effect involves the computation of a propagator for the incident particles and the result, as explained by Wu and Yang,[2] depends only on the loop variable $\exp[-\oint A_\mu\, dx^\mu]$. This has been elaborated extensively by Horvathy.[3] In Ref. 3, a gauge analogous to the U gauge in monopole theory is chosen so that the Yang–Mills potential is diagonal in the internal SU(2) group space in order to facilitate the calculation. It is well-known that because of the impenetrable solenoid (we use the term "solenoid" to indicate the source for the gauge fields, whether Abelian or non-Abelian), the propagator can be expressed as a sum of the free homotopy propagators weighted by the nonintegrable phase factors corresponding to homotopically distinct paths.[4,5] The purpose of this paper is to generalize the work of Refs. 4 and 5 to the non-Abelian case and provide a complementary approach to that of Ref. 3. We also obtain an explicit expression for the propagator when there are two solenoids present.

In an idealized setup involving only a single solenoid, the incident test particles are confined to a gauge curvature-free region M, which can be taken as a punctured plane $R^2-\{0\}$ and is multiply connected. Inside the infinitely long solenoid located at the origin there resides a strong non-Abelian gauge field strength $F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu+[A_\mu,A_\nu]$. The multiple connectedness of M is necessary for observing the AB effect since every flat connection (that is, with null curvature) defined on a simply connected manifold is trivial.[6] In other words, on a nonsimply connected space a nontrivial gauge field may exist with a vanishing field strength. One can deal with a multiply connected space M through its covering space, in particular the universal covering space $\widetilde{M}$.[4,5] One has $M=\widetilde{M}/\Gamma$, where Γ is a discrete group of diffeomorphisms of $\widetilde{M}$ and has no fixed point. Single-valued functions on M can then be lifted to functions constant on fibers in $\widetilde{M}$. By lifting paths on M to $\widetilde{M}$, we shall see that the lift of a nontrivial connection on M is trivial on $\widetilde{M}$.

It should be noted that a gauge transformation $\widetilde{U}(\tilde{x})$ of $\widetilde{A}_\mu(\tilde{x})=0$ on the universal covering space $\widetilde{M}$ can descend to M in three different ways.

(i) The first way is when $\widetilde{U}(\tilde{x})$ is constant on the fibers and can therefore be projectable to be a well-defined single-valued gauge transformation on M. This corresponds to no AB effect.

(ii) The second way is when $\widetilde{U}(\tilde{x})$ is not projectable, but nevertheless gives rise to a projectable gauge field $\widetilde{A}_\mu(\tilde{x})=\widetilde{U}\tilde{\partial}_\mu\widetilde{U}^{-1}$. An example is that $\widetilde{U}(\tilde{x})$ is quasiperiodic on the fibers. This may result in the AB effect on M. Conversely, a gauge field producing the AB effect on M can be lifted to a trivial gauge field $\widetilde{A}_\mu=\widetilde{U}\tilde{\partial}_\mu\widetilde{U}^{-1}$, but with a nonprojectable $\widetilde{U}$. The fact that the gauge transformation $\widetilde{U}(\tilde{x})$ in the universal covering space $\widetilde{M}$ is not constant on its fibers results in a nontrivial phase factor connecting any two different points of any single fiber, which is necessary for observing the AB effect.

(iii) The third way is when $\widetilde{U}(\tilde{x})$ is a nonprojectable gauge transformation and the resulting pure gauge $\widetilde{A}_\mu=\widetilde{U}\tilde{\partial}_\mu\widetilde{U}^{-1}$ is also not projectable.

II. PATH-INDEPENDENT PHASE FACTORS IN THE COVERING SPACE

In the region M outside the solenoid, the gauge field potential is well defined and locally can be written in the pure gauge form

$$A_\mu=U\partial_\mu U^{-1}\tag{1}$$

since in M, $F_{\mu\nu}=0$. Expression (1) implies that the gauge transformation $U(x)$ can be written as a non-Abelian phase factor

$$U(x)=\exp\left(-\int_{x_{\rm ref}}^{x}A_\mu\, dx^\mu\right),\tag{2}$$

where the path ordering is always understood and $x_{\rm ref}$ is a fixed reference point. For the AB effect there is a strong flux through the origin of M and, consequently, it is *not* true that along any closed loop c,

 0022-2488/88/051154-04$02.50

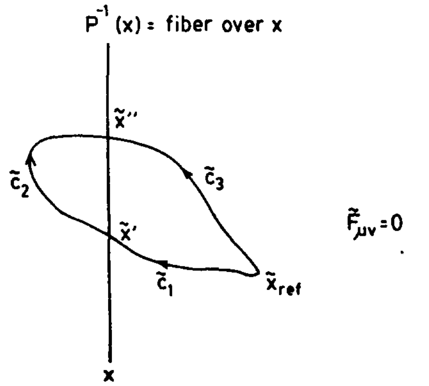


FIG. 1. This figure is used to establish Eq. (10). Here $\tilde{x}'$ and $\tilde{x}''$ are on the fiber over x and $\tilde{c}_i$, $i = 1,2,3$, are paths in $\widetilde{M}$, where $\widetilde{F}_{\mu\nu} = 0$.

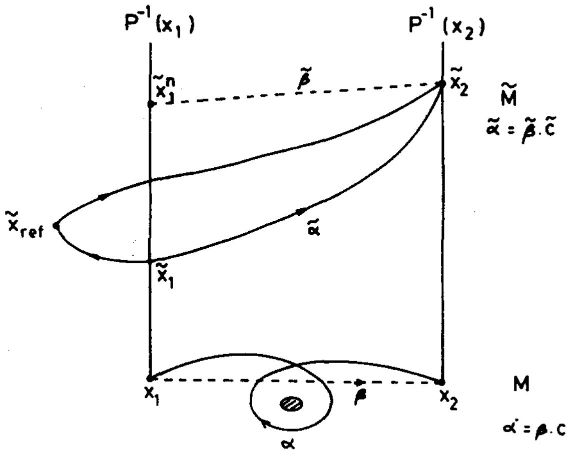


FIG. 2. The derivation of Eq. (14) is illustrated. The location of the point $\tilde{x}_1''$ on the fiber over x_1 determines the homotopy class $[c]$ of the closed path c in M.

$$\exp - \oint_c A_\mu \, dx^\mu = I(\text{identity}). \tag{3}$$

This prevents us from writing A_μ globally as a pure gauge form. However, A_μ can still be written as in expression (1) everywhere on M except along a line extending from the origin to infinity. In this case, the gauge transformation is singular along that line. This does not mean that A_μ is singular on the half-line; instead it just means that A_μ cannot be written as a pure gauge form along the half-line. It must be stressed that A_μ is nonsingular everywhere on M.

Since A_μ and $F_{\mu\nu}$ are regular everywhere on M, they can be lifted to be universal covering space $\widetilde{M}$ of M:

$$\widetilde{F}_{\mu\nu}(\tilde{x}) = F_{\mu\nu}(x) = 0, \tag{4}$$

$$\widetilde{A}_\mu(\tilde{x}) = A_\mu(x). \tag{5}$$

Here x is the projection of $\tilde{x}$, $P(\tilde{x}) = x$, or $\tilde{x}$ is any point on the fiber over $x \in M$. As $\widetilde{M}$ is simply connected and $\widetilde{F}_{\mu\nu} = 0$ everywhere on $\widetilde{M}$, the lifted gauge field $\widetilde{A}_\mu$ can be written as a pure gauge form globally:

$$\widetilde{A}_\mu(\tilde{x}) = \widetilde{U}(\tilde{x})\tilde{\partial}_\mu \widetilde{U}(\tilde{x})^{-1}, \tag{6}$$

with

$$\widetilde{U}(\tilde{x}) = \exp\left(- \int_{\tilde{x}_{ref}}^{\tilde{x}} \widetilde{A}_\mu(\tilde{x}) d\tilde{x}^\mu \right). \tag{7}$$

By the non-Abelian Stokes theorem,[7] $\widetilde{U}(\tilde{x})$ is path independent and single valued on $\widetilde{M}$:

$$\exp\left(- \oint_{\tilde{c}} \widetilde{A}_\mu \, d\tilde{x}^\mu \right) = I, \tag{8}$$

where $\tilde{c}$ is any closed path in $\widetilde{M}$.

Note that $\widetilde{A}_\mu(\tilde{x})$ is constant on the fiber over any $x \in \widetilde{M}$, that is, it is projectable to a single-valued potential $A_\mu(x)$ on M although $\widetilde{U}(\tilde{x})$, which is defined only on $\widetilde{M}$, needs not be constant on the fiber. Indeed, for any two preimage points, $\tilde{x}', \tilde{x}'' \in P^{-1}(\tilde{x})$, $\widetilde{U}(\tilde{x}')$, and $\widetilde{U}(\tilde{x}'')$ have a simple relation. On $\widetilde{M}$ and using Fig. 1, it is easy to see that

$$\exp\left(- \int_{\tilde{x}'}^{\tilde{x}''} \widetilde{A}_\mu \, dx^\mu \Big|_{\tilde{c}_2} \right) \cdot \exp\left(- \int_{\tilde{x}_{ref}}^{\tilde{x}'} \widetilde{A}_\mu \, d\tilde{x}^\mu \Big|_{\tilde{c}_1} \right)$$
$$= \exp\left(- \int_{\tilde{x}_{ref}}^{\tilde{x}''} \widetilde{A}_\mu \, d\tilde{x}^\mu \Big|_{\tilde{c}_3} \right). \tag{9}$$

The particular paths $\tilde{c}_i$ are not important since Eq. (8) is satisfied on $\widetilde{M}$. From Eq. (7), it follows that

$$\widetilde{U}(\tilde{x}'')\widetilde{U}^{-1}(\tilde{x}') = \exp\left(- \int_{\tilde{x}'}^{\tilde{x}''} \widetilde{A}_\mu \, d\tilde{x}^\mu \Big|_{\tilde{c}_2} \right). \tag{10}$$

Since $\tilde{x}', \tilde{x}'' \in P^{-1}(x)$, it is clear that $P(\tilde{c}_2)$ is a closed path c in M, although $\tilde{c}_2$ is not closed in $\widetilde{M}$. As $\widetilde{A}_\mu(\tilde{x}) = A_\mu(x)$, we can write

$$\exp\left(- \int_{\tilde{x}'}^{\tilde{x}''} \widetilde{A}_\mu \, d\tilde{x}^\mu \right) = \exp\left(- \oint_c A_\mu \, dx^\mu \right), \tag{11}$$

which depends only on the homotopy class $[c]$ of the path c, where $[c]$ is an element of the fundamental group $\pi_1(M)$ of M. Note that $\pi_1(M)$ is isomorphic to the discrete group Γ of diffeomorphisms of M.

III. THE PROPAGATOR

We proceed to calculate the nonrelativistic propagator $K(x_2,t_2;x_1,t_1)$, for the AB effect in M:

$$\psi(x_2,t_2) = \int K(x_2,t_2;x_1,t_1)\psi(x_1,t_1)dx_1. \tag{12}$$

Consider a partial propagator for an arbitrary path α on M: $K_\alpha(x_2,t_2;x_1,t_1)$. The path α can be decomposed as $\beta \cdot c$, where β is a reference path in M and c is a closed path. Assigning a particular preimage $\tilde{x}_1$ to x_1, there is a unique lifted path[8] $\tilde{\alpha} = \tilde{\beta}\cdot\tilde{c}$ from $\tilde{x}_1$ to $\tilde{x}_2$. The partial propagator K_α on M can then be lifted to the covering space such that

$$K_\alpha(x_2,t_2;x_1,t_1) = \widetilde{K}_{\tilde{\alpha}}(\tilde{x}_2,t_2;\tilde{x}_1 t_1). \tag{13}$$

Since $\widetilde{A}_\mu$ as given by Eq. (6) is a gauge transform of $\widetilde{A}_\mu = 0$ on $\widetilde{M}$, $\widetilde{K}_{\tilde{\alpha}}$ is simply the gauge transform of the free propagator $K^0_{\tilde{\alpha}}$ and may be evaluated with the help of Fig. 2. One has

$$
\begin{aligned}
\tilde{K}_{\tilde{\alpha}}(\tilde{x}_2,t_2;\tilde{x}_1,t_1) &= \tilde{U}(\tilde{x}_2)\tilde{K}^0_{\tilde{\alpha}}(\tilde{x}_2,t_2;\tilde{x}_1,t_1)\tilde{U}^{-1}(\tilde{x}_1) \\
&= \left[\exp\left(-\int_{\tilde{x}_{\rm ref}}^{\tilde{x}_2}\tilde{A}_\mu\, d\tilde{x}^\mu\right)\right]\tilde{K}^0_{\tilde{\alpha}}(\tilde{x}_2,t_2;\tilde{x}_1,t_1)\left[\exp\left(-\int_{\tilde{x}_{\rm ref}}^{\tilde{x}_1}\tilde{A}_\mu\, d\tilde{x}^\mu\right)\right]^{-1} \\
&= \left[\exp\left(-\int_{\tilde{x}_1}^{\tilde{x}_2}\tilde{A}_\mu\, d\tilde{x}^\mu\Big|_{\tilde{\alpha}}\right)\right]\tilde{K}^0_{\tilde{\alpha}}(\tilde{x}_2,t_2;\tilde{x}_1,t_1) \\
&= \left[\exp\left(-\int_{\tilde{x}_1^n}^{\tilde{x}_2}\tilde{A}_\mu\, d\tilde{x}^\mu\Big|_{\tilde{\beta}}\right)\right]\left[\exp\left(-\int_{\tilde{x}_1}^{\tilde{x}_1^n}\tilde{A}_\mu\, d\tilde{x}^\mu\Big|_{\tilde{c}}\right)\right]\tilde{K}^0_{\tilde{\alpha}}(\tilde{x}_2,t_2;\tilde{x}_1,t_1) \\
&= \left[\exp\left(-\int_{x_1}^{x_2}A_\mu\, dx^\mu\Big|_{\beta}\right)\right]\left[\exp\left(-\oint_c A_\mu\, dx^\mu\right)\right]K^0_\alpha(x_2,t_2;x_1,t_1). \qquad (14)
\end{aligned}
$$

The point $\tilde{x}_1^n$ in $\tilde{M}$ is on the fiber over x_1 and its location on the fiber determines the homotopy class $[c]$. Writing the nonintegrable phase factor along the reference path β as B, we obtain, from Eqs. (13) and (14),

$$
\begin{aligned}
K(x_2,t_2;x_1,t_1) &= \sum_{[c]\in\pi_1(M)}\sum_{\substack{\alpha=\beta\cdot c\\ c\in[c]}} K_\alpha(x_2,t_2;x_1,t_1) \\
&= B\sum_{[c]\in\pi_1(M)}\sum_{\substack{\alpha=\beta\cdot c\\ c\in[c]}}\left[\exp\left(-\oint_c A_\mu\, dx^\mu\right)\right] \\
&\quad\times K^0_\alpha(x_2,t_2;x_1,t_1). \qquad (15)
\end{aligned}
$$

The first summation over paths α is performed by summing over all closed paths c within a homotopy class $[c]$; this is then followed by the summation over all homotopy classes of $\pi_1(M)$. The loop variable $\exp(-\oint_c A_\mu\, dx^\mu)$ depends only on the homotopy class $[c]$. Factoring out the free propagator K^0_α along the reference path β, the final result is

$$
\begin{aligned}
K(x_2,t_2;x_1,t_1) &= B'\sum_{[c]\in\pi_1(M)}\left[\exp\left(-\oint_{[c]}A_\mu\, dx^\mu\right)\right] \\
&\quad\times\sum_{c\in[c]}K^0_c(x_2,t_2;x_1,t_1), \qquad (16)
\end{aligned}
$$

where B' is the product of B and the free propagator along the reference path β. Note that in the above discussion M needs not be $R^2-\{0\}$, as indicated in Fig. 2; the result (16) is in fact valid for any multiply connected space M; and $\pi_1(M)$ may be non-Abelian. For the single solenoid, $\pi_1(M)$ is Abelian, its elements are labeled by integer n, and Eq. (16) can be rewritten as

$$
K(x_2,t_2;x_1,t_1) = B'\sum_n \Phi^n K^0_n(x_2,t_2;x_1,t_1), \qquad (17)
$$

where we have defined

$$
K^0_n(x_2,t_2;x_1t_1) = \sum_{c\in[n]}K^0_c(x_2,t_2;x_1,t_1) \qquad (18)
$$

and Φ denotes the closed loop phase factor when $n=1$. Equation (17) is same as that given by Refs. 3–5. The loop variables $\exp(-\oint_c A_\mu\, dx^\mu)$ are elements of the holonomy group from $\pi_1(M)$ to the gauge group.

In summary, the partial propagator along a path in the multiply connected space M is lifted to the universal covering space $\tilde{M}$, where it can be regarded as a gauge transform of a free partial propagator. The phase factor in $\tilde{M}$ is path independent since the field strength $\tilde{F}_{\mu\nu}$ vanishes everywhere and $\tilde{M}$ is simply connected. An open path (chain) in $\tilde{M}$ with two end points on the same fiber is projected to M as a closed path (cycle). Thus the loop variable $\exp(-\oint A_\mu\, dx^\mu)$ emerges naturally in expression (16). Because this loop variable assumes different values for different homotopy classes due to the multiple connectedness of M, we have the AB effect. We note from Eqs. (10) and (11) that there will be no AB effect precisely when $\tilde{U}(\tilde{x}'') = \tilde{U}(\tilde{x}')$, that is, if $\tilde{U}(\tilde{x})$ is constant on fibers and hence projectable.

IV. TWO OR MORE SOLENOIDS

The result (16) is valid for any multiply connected space M and any gauge group G, although, in our derivation, we have made use of Fig. 2, which tends to suggest that M is $R^2-\{0\}$ and $\pi_1(M)$ is Abelian. We now consider the case of k solenoids; M is then equivalent topologically to a plane with k disks removed. For the case $k=2$, M is equivalent to a figure eight. The fundamental group $\pi_1(M)$ is non-Abelian and is the free product of $Z*Z*\cdots*Z$ of k copies of the infinite cyclic group.[8] As an illustration for the non-Abelian $\pi_1(M)$, we discuss the two-solenoid case. Every closed loop in M is homotopic to the combination $c_1^n\cdot c_2^m$ or $c_2^n\cdot c_1^m$, where c_1 and c_2 are closed loops going once around the solenoids 1 and 2, respectively. Thus from Eq. (16) we obtain

$$
\begin{aligned}
K(x_2,t_2;x_1t_1) &= B'\sum_{n,m}\sum_{\substack{i,j=1\\ i\neq j}}^{2}\left[\exp\left(-\oint_{[c_i]}A_\mu\, dx^\mu\right)\right]^n \\
&\quad\times\left[\exp\left(\oint_{[c_j]}A_\mu dx^\mu\right)^m\right] \\
&\quad\cdot\sum_{\gamma\in[c_i^n\cdot c_j^m]}K^0_\gamma(x_2,t_2;x_1,t_1). \qquad (19)
\end{aligned}
$$

Note that knowledge of the fundamental group $\pi_1(M)$ is needed in order to write down the more explicit propagator (19) from Eq. (16).

V. COMMENTS

We now proceed to make a few remarks.

(i) On substituting Eq. (16) into Eq. (12), we immediately have

$$\psi(x_2,t_2) = B' \sum_{[c]\in\pi_1(M)} \left[\exp\left(-\oint A_\mu\, dx^\mu\right)\right]\psi^0_{[c]}(x_2,t_2), \tag{20}$$

where

$$\psi^0_{[c]}(x_2,t_2) \equiv \int \sum_{c\in[c]} K^0_c(x_2,t_2;x_1,t_1)\psi(x_1,t_1)dx_1. \tag{21}$$

Expressions (20) and (21) can also be written in the covering space $\widetilde{M}$. Expression (20) has appeared in the literature before[9] and indicates that the wave function at the detector in the generalized AB experiment is the sum of "partial amplitudes" weighted by different nonintegrable phase factors.

(ii) Surveying the literature, we found that Eq. (16) has also been discussed in Ref. 10. However, the starting point of Ref. 10 is to assume that the propagator is given by

$$K(x_2,y_2;x_1,t_1) = \int_{x_1(t_1)}^{x_2(t_2)} \left(\exp \int A_i\, dx^i \Big|_\Gamma\right)\exp[iS(\Gamma)],$$

where $S(\Gamma)$ is the free action, whereas in our approach we start by lifting a partial propagator along a path to the universal covering space.

(iii) The Wong equation[11] describes the motion of a particle with an internal non-Abelian charge interacting with an external Yang–Mills field

$$(\partial_\mu \mathbf{I} + \mathrm{ad}(A_\mu)\mathbf{I})\frac{dx^\mu}{d\tau} = 0, \tag{22}$$

where $\mathbf{I}$ is the non-Abelian charge vector. On the space M, one can solve Eq. (22) by parallel transport:

$$\mathbf{I}(x) = \mathrm{ad}\left[\exp\left(-\int_{x_0}^{x} A_\mu\, dx^\mu\Big|_\alpha\right)\right]\mathbf{I}(x_0), \tag{23}$$

where $\mathbf{I}(x)$ is dependent on the path from x_0 to x. For paths α_1 and α_2 such that $\alpha_2^{-1}\alpha_1$ is a closed path in M, the internal vector $\mathbf{I}$ transported along paths α_1 and α_2 differ by the Wu–Yang factor, a fact used in Refs. 3 and 12 to discuss the non-Abelian AB effect. Equation (23) can also be obtained in the universal covering space $\widetilde{M}$. On $\widetilde{M}$ the gauge field is a pure gauge everywhere, $\widetilde{A}_\mu = \widetilde{U}\widetilde{\partial}_\mu\widetilde{U}^{-1}$, and the Wong equation becomes

$$(\widetilde{\partial}_\mu \mathbf{I} + \widetilde{U}\widetilde{\partial}_\mu\widetilde{U}^{-1}\mathbf{I})\frac{dx^\mu}{d\tau} = 0. \tag{24}$$

Equation (24) can be simplified to

$$\widetilde{\partial}_\mu(\widetilde{U}^{-1}\mathbf{I}(\tilde{x})) = 0,$$

giving

$$\mathbf{I}(\tilde{x}) = \mathrm{ad}(\widetilde{U}(\tilde{x})\widetilde{U}^{-1}(\tilde{x}_0))\mathbf{I}(\tilde{x}_0),$$

which is the same as Eq. (23).

(iv) From expression (16), we confirm the viewpoint of Ref. 2 that the classes of nonintegrable phase factor provide a complete description of the physical situation. Consider two gauge fields A_μ and A'_μ in M, with the corresponding lifts $\widetilde{A}_\mu = \widetilde{U}\ \widetilde{\partial}_\mu\widetilde{U}^{-1}$ and $\widetilde{A}'_\mu = \widetilde{U}'\widetilde{\partial}_\mu\widetilde{U}'^{-1}$ in $\widetilde{M}$; the fields are gauge transformable to each other in M iff $\tilde{g}(\tilde{x}) = \widetilde{U}'(\tilde{x})\widetilde{U}^{-1}(\tilde{x})$ is projectable. One can verify, using Eqs. (10) and (11), that this happens if the nonintegrable phase factors $\exp(-\oint A_\mu\, dx^\mu)$ and $\exp(-\oint A'_\mu\, dx^\mu)$ belong to the same conjugate class. Hence no experiment in M can differentiate between the two nonintegrable phase factors.

[1]Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).

[2]T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845 (1975).

[3]P. A. Horvathy, Phys. Rev. D **33**, 407 (1986); "The Wu–Yang factor and the non-Abelian Aharonov–Bohm experiment," preprint DIAS-STP-86-04, 1986.

[4]L. S. Schulman, J. Math. Phys. **12**, 304 (1971).

[5]J. S. Dowker, J. Phys. A **5**, 936 (1972).

[6]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley–Interscience, New York, 1963), Vol. I.

[7]P. Goddard and D. Olive, Rep. Prog. Phys. **41**, 1357 (1978).

[8]W. S. Massey, *Algebraic Topology: An Introduction* (Harcourt, Brace and World, New York, 1967); M. A. Armstrong, *Basic Topology* (McGraw-Hill, New York, 1979).

[9]C. C. Bernido and A. Inomata, J. Math. Phys. **22**, 715 (1981).

[10]R. Sundrum and L. J. Tassie, J. Math. Phys. **27**, 1566 (1986).

[11]S. K. Wong, Nuovo Cimento **65A**, 689 (1970).

[12]L. C. L. Botelho and J. C. deMello, J. Phys. A **18**, 2633 (1985).

Symmetry properties of product states for the system of N n-level atoms

Robert Alicki, Sławomir Rudnicki, and Sławomir Sadowski
Institute of Theoretical Physics and Astrophysics, University of Gdańsk, PL-80-952 Gdańsk, Poland

(Received 16 July 1987; accepted for publication 16 December 1987)

Suppose that the state of a system of N n-level atoms is given by a tensor product of N identical density matrices. The exact formulas are presented that describe the probability that such a system may be found in a pure state with a given symmetry with respect to permutations of atoms. The asymptotic form of these probabilities valid for large N is also derived.

I. INTRODUCTION

The origin of the problem solved in the present paper may be found in the theory of collective phenomena in quantum optics like superradiance,[1–3] subradiance,[2–4] and limited thermalization.[5] In the simplest mean-field model used in this theory the interaction of the system of N n-level atoms with an electromagnetic field is invariant with respect to permutation of atoms. On the other hand the N-atomic states span the whole N-fold tensor product Hilbert space because the spatially separated atoms are treated as distinguishable objects.[1] Hence in principle the states with arbitrary symmetry with respect to permutation of atoms may occur.[3,4] As a consequence the probability that the atomic system may be found in a state with a given symmetry with respect to permutations is a constant of motion. This is a very strong restriction on the time evolution of the system. One can use this property to estimate the energy emitted in a superradiance pulse for different initial conditions[4] and to describe the form of the corresponding final states (subradiance and limited thermalization[4,5]) in terms of the probabilities mentioned above. We shall show in the present paper that these probabilities may be exactly calculated for the initial states being tensor products of N identical density matrices. Such states are often proposed as initial states of the atomic system.[3,4] The obtained formulas are rather complicated and therefore the asymptotic expressions valid for $N\to\infty$ will be derived also. One should mention that the discussed problem was formulated and partially solved in Ref. 4 for the case of two-level atoms.

II. MAIN RESULTS

The Hilbert space for a single atom is denoted by $\mathscr{H}=\mathbb{C}^n$ while the Hilbert space of the N-atom system is $\mathscr{H}^{(N)}=\otimes_N\mathscr{H}=\mathbb{C}^{(n^N)}$. By $\rho^{(N)}$ we denote a state of the atomic system given by a tensor product

$$\rho^{(N)}=\bigotimes_N \rho. \tag{2.1}$$

We choose in $\mathscr{H}$ an orthonormal basis $\{|k\rangle;\ k=1,2,...,n\}$ which diagonalizes the density matrix ρ,

$$\rho=\sum_{k=1}^{n}\rho_k|k\rangle\langle k|. \tag{2.2}$$

Now, we recall some standard definitions and results of the group theory.[6–8]

By $\Lambda^{(N)}$ we denote the set of Young frames $\lambda=(\lambda_1,...,\lambda_n)$, $\lambda_k=0,1,...;\ \lambda_k\geqslant\lambda_{k+1};\ \Sigma_{k=1}^n\lambda_k=N$. For any standard Young tableau $T_\lambda^{(\alpha)}$ (α labels different standard tableaux associated with λ) we have an idempotent operator $e(T_\lambda^{(\alpha)})$ projecting on the subspace $L_\lambda^{(\alpha)}=e(T_\lambda^{(\alpha)})\mathscr{H}^{(N)}$ of the "tensors of a given symmetry type." Here $\{L_\lambda^{(\alpha)}\}$ are carrier spaces for the irreducible representations of GL(n), U(n), SU(n), which are equivalent for a fixed λ. Because $L_\lambda^{(\alpha)}$, $L_\lambda^{(\beta)}$ are generally not orthogonal one may replace them by the orthogonal subspaces $\mathscr{H}_\lambda^{(\alpha)}$, which remain the carrier spaces for GL(n), etc.

Our problem is to find the probabilities $\mathrm{tr}(\rho^{(N)}Q_\lambda^{(\alpha)})$, $\mathrm{tr}(\rho^{(N)}P_\lambda^{(\alpha)})$, $\mathrm{tr}(\rho^{(N)}P_\lambda)$ where $Q_\lambda^{(\alpha)}$, $P_\lambda^{(\alpha)}$, P_λ are orthogonal projectors on $L_\lambda^{(\alpha)}$, $\mathscr{H}_\lambda^{(\alpha)}$, and $\oplus_\alpha\mathscr{H}_\lambda^{(\alpha)}$, respectively. Due to the invariance of a trace to a similarity transformation one obtains

$$\mathrm{tr}(\rho^{(N)}Q_\lambda^{(\alpha)})=\mathrm{tr}(\rho^{(N)}P_\lambda^{(\alpha)})=\mathscr{N}(\lambda)^{-1}\,\mathrm{tr}(\rho^{(N)}P_\lambda), \tag{2.3}$$

where $\mathscr{N}(\lambda)$ is a number of standard tableaux for a fixed λ and is given by (see Ref. 7, p. 191, Ref. 8, p. 123)

$$\mathscr{N}(\lambda)=\frac{N!\Pi_{i<j}^n(\nu_i-\nu_j)}{\nu_1!\nu_2!\cdots\nu_n!},\quad \nu_j=\lambda_j+n-j. \tag{2.4}$$

Assume first that ρ is not strictly positive and hence say $\rho_{r+1}=\rho_{r+2}=\cdots=\rho_n=0$ for a fixed r, $0<r<n-1$. Therefore the searched probabilities may be different from 0 only if $\lambda=(\lambda_1,...,\lambda_r,0,...,0)$. Moreover the nonzero contribution comes from the vectors of the type

$$|\psi\rangle=\sum_{k_1,...,k_N=1}^{n}C_{k_1,...,k_N}|k_1,...,k_N\rangle \tag{2.5}$$

with the coefficients satisfying the condition $C_{k_1,...,k_N}=0$ if any $k_s=r+1,r+2,...,n$, $s=1,2,...,N$. It follows that the problem can be reduced to the lower-dimensional case with $\mathscr{H}=\mathbb{C}^r$ and hence we assume always that

$$\rho_k>0,\quad k=1,2,...,n. \tag{2.6}$$

For a given set of eigenvalues $\rho_1,\rho_2,...,\rho_n$ we define the sequence of natural numbers $\mu_1,\mu_2,...,\mu_n$. Let ρ_k be a degenerated eigenvalue with the multiplicity d_k and let the value of ρ_k appear w_k times in the sequence $\rho_1,\rho_2,...,\rho_k$. Then we set

$$\mu_k=d_k-w_k\,. \tag{2.7}$$

Now we are able to formulate the first theorem.

Theorem 1: With the notation and asumptions as above

 0022-2488/88/051158-05$02.50

$$\text{(i) } \mathrm{tr}\Big(\underset{N}{\otimes}\rho\, Q_\lambda^{(\alpha)}\Big) = \mathrm{tr}\Big(\underset{N}{\otimes}\rho\, P_\lambda^{(\alpha)}\Big) = \det\Delta/\det\Gamma \quad (2.8)$$

and is independent from α,

$$\text{(ii) } \mathrm{tr}\Big(\underset{N}{\otimes}\rho\, P_\lambda\Big) = \mathrm{N}(\lambda)\det\Delta/\det\Gamma, \quad (2.9)$$

where $\Delta = (\Delta_{kl})$, $\Gamma = (\Gamma_{kl})$, $k,l = 1,2,\ldots,n$,

$$\Delta_{kl} = \begin{cases} \rho_l^{\nu_k}, & \text{for } \mu_l = 0, \\ \nu_k(\nu_k-1)\cdots(\nu_k-\mu_l+1)\rho_l^{\nu_k-\mu_l}, & \text{for } \mu_l > 0, \end{cases} \quad (2.10)$$

$$\Gamma_{kl} = \begin{cases} \rho_l^{n-k}, & \text{for } \mu_l = 0, \\ (n-k)(n-k-1)\cdots(n-k-\mu_l+1)\rho_l^{n-k-\mu_l}, & \text{for } \mu_l > 0. \end{cases} \quad (2.11)$$

Remark: For the density matrix ρ with nondegenerated eigenvalues $\Delta_{kl} = \rho_l^{\lambda_k+n-k}$ and $\det\Gamma$ is the so-called Vandermonde determinant equal to

$$\det\Gamma = \prod_{i<j=2}^{n} (\rho_i - \rho_j)\,. \quad (2.12)$$

Here $\mathscr{P}(\lambda) = \mathrm{tr}(\otimes_N \rho\, P_\lambda)$ defines a probability distribution on the set $\Lambda^{(N)}$ of Young tableaux. The formula (2.9) is rather complicated and not very transparent. Therefore it is worthwhile to derive a simpler asymptotic formula valid for $N \to \infty$.

We introduce the following notation: (i) $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_n$ is the reordered set of the eigenvalues $\rho_1,\rho_2,\ldots,\rho_n$;

$$\text{(ii) } \xi_k^{(N)} = \frac{\lambda_k - Np_k}{(Np_k(1-p_k))^{1/2}}, \quad k = 1,2,\ldots,n-1, \quad (2.13)$$

denote $n-1$ standardized random variables on $\Lambda^{(N)}$, (iii) the degeneration of $p_1,p_2,\ldots,p_n$ is taken into account by introducing $\{\bar p_\alpha\}$ and $\{d_\alpha\}$, $\alpha = 1,2,\ldots,s \leqslant n$, such that

$$\begin{aligned} \bar p_1 &= p_1 = \cdots = p_{d_1}, \\ \bar p_2 &= p_{d_1+1} = \cdots = p_{d_1+d_2}, \\ &\vdots \\ \bar p_s &= p_{d_1+d_2+\cdots+d_{s-1}+1} = \cdots = p_n; \end{aligned} \quad (2.14)$$

and (iv) D_α is a subset of consecutive natural numbers,

$$\begin{aligned} D_1 &= \{1,2,\ldots,d_1\}, \\ D_\alpha &= \{(d_1+d_2+\cdots+d_{\alpha-1}+1),\ldots, \\ &\quad (d_1+d_2+\cdots+d_\alpha)\}, \end{aligned} \quad (2.15)$$

for $\alpha = 2,3,\ldots,s$.

Theorem 2: Let $F^{(N)}(x_1,x_2,\ldots,x_{n-1})$ denote the probability that

$$\xi_1^{(N)} \leqslant x_1,\ \xi_2^{(N)} \leqslant x_2,\ldots,\xi_{n-1}^{(N)} \leqslant x_{n-1}\,.$$

Then

$$\begin{aligned} &\lim_{N\to\infty} F^{(N)}(x_1,x_2,\ldots,x_{n-1}) \\ &= \mathscr{N}\int_{-\infty}^{x_1} dy_1 \int_{-\infty}^{x_2} dy_2 \cdots \int_{-\infty}^{x_{n-1}} dy_{n-1} \\ &\quad \times \chi_\Omega(y_1,y_2,\ldots,y_{n-1}) \prod_{\alpha=1}^{s} \prod_{\substack{1\leqslant i<j\leqslant n \\ i,j\in D_\alpha}} (y_i - y_j)^2 \\ &\quad \times \exp\Big\{-\frac{1}{2}\sum_{k,l=1}^{n-1} A_{kl}\, y_k y_l\Big\}, \end{aligned} \quad (2.16)$$

where (a) $\chi_\Omega(\cdot)$ is a characteristic function of the following subset of $\mathbf{R}^{n-1}$:

$$\begin{aligned} \Omega = \Big\{&(y_1,y_2,\cdots,y_{n-1}); y_1\in\mathbf{R};\ y_k\in\mathbf{R} \text{ if } p_k < p_{k-1} \\ &\text{or } y_k \leqslant y_{k-1} \text{ if } p_k = p_{k-1} \text{ for } k = 2,\ldots,n, \\ &\text{and } y_n = -\sum_{i=1}^{n-1} y_i\Big(\frac{p_i(1-p_i)}{p_n(1-p_n)}\Big)^{1/2}\Big\}; \end{aligned}$$

$$\text{(b) } A_{kl} = (\delta_{kl} + p_n^{-1}(p_k p_l)^{1/2})(1-p_k)^{1/2}(1-p_l)^{1/2}, \quad k,l = 1,2,\ldots,n-1; \quad (2.17)$$

$$\begin{aligned} \text{(c) } \mathscr{N} &= (2\pi)^{-(n-1)/2}(\det\Gamma)^{-1}p_n^{1/2} \\ &\quad \times \prod_{k=1}^{n-1}(1-p_k)^{-1/2} \prod_{\alpha<\beta}(\bar p_\alpha - \bar p_\beta)^{d_\alpha d_\beta} \\ &\quad \times \prod_\alpha (1-\bar p_\alpha)^{d_\alpha(d_\alpha-1)/2}. \end{aligned} \quad (2.18)$$

Remarks: The heuristic meaning of Theorem 2 is the following: For large N and in the case of nondegenerated eigenvalues $\{\rho_k\}$ the probability distribution $\mathscr{P}(\lambda)$ has a sharp maximum roughly corresponding to $\bar\lambda \cong (Np_1, Np_2,\ldots,Np_n)$, which lies in the interior of $\Lambda^{(N)}$ and the Gaussian shape with a width proportional to $N^{1/2}$. For degenerated eigenvalues the point $\bar\lambda$ lies on the boundary of $\Lambda^{(N)}$, the maximum of $\mathscr{P}(\lambda)$ is shifted into the interior at the distance proportional to $N^{1/2}$, and the shape of $\mathscr{P}(\lambda)$ is modified by the presence of the polynomial term.

III. PROOFS OF THE THEOREMS

We use the definitions and notation introduced in the previous section.

It follows from (2.3) that it is enough to calculate $\mathrm{tr}(\otimes_N\rho\, Q_\lambda^{(\alpha)})$ for arbitrary α, say $\alpha = 1$. The subspace $L_\lambda^{(1)}$ may be constructed as follows. Let $\mathscr{K}_\lambda^{(1)}$ denote a set of standard sequences $K = (k_1,k_2,\ldots,k_N)$ such that $k_j = 1,2,\ldots,n$ are distributed in the Young frame in nondecreasing order in every row from left to right and in increasing order in every column from top to bottom and the indices of $\{k_j\}$ form the standard tableau $T_\lambda^{(1)}$. The subspace $L_\lambda^{(1)}$ is spanned by the linearly independent vectors of the form[6–8]

$$e(T_\alpha^{(1)})|k_1,k_2,\ldots,k_N\rangle. \quad (3.1)$$

The vector $|k_1,k_2,...,k_N\rangle$ corresponds to a standard sequence $K=(k_1,k_2,...,k_N)\in\mathscr{K}_\lambda^{(1)}$. For any $K\in\mathscr{K}_\lambda^{(1)}$ we define a sequence of natural numbers $N_1^{(K)}, N_2^{(K)},...,N_n^{(K)}$ such that $N_r^{(K)}$ describes how many times the number $k_j=r$ appears in the sequence $K=(k_1,k_2,...,k_N)$.

The subspace $L_\lambda^{(1)}$ may be decomposed into an orthogonal sum of subspaces

$$L_\lambda^{(1)}=\bigoplus_{\substack{(N_1,N_2,...,N_n)\\ \Sigma_k N_k=N}} L_\lambda^{(1)}(N_1,N_2,...,N_n), \tag{3.2}$$

where $L_\lambda^{(1)}(N_1,N_2,...,N_n)$ is spanned by the vectors given by Eq. (3.1) with $N_j^{(K)}=N_j$. Obviously for $|\varphi\rangle\in L_\lambda^{(1)}(N_1,N_2,...,N_n)$,

$$\otimes_N\rho|\varphi\rangle=\rho_1^{N_1}\rho_2^{N_2}\cdots\rho_n^{N_n}|\varphi\rangle\,. \tag{3.3}$$

Lemma: Let ρ be a Hermitian matrix on $\mathbb{C}^n$ with nondegenerated eigenvalues $0<\rho_k<1$, $k=1,2,...,n$. Then

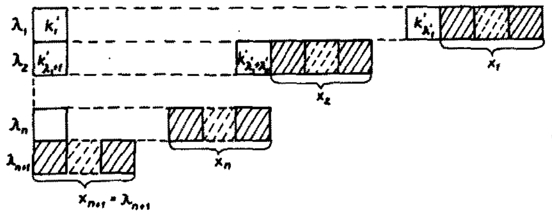
$$\mathrm{tr}(\otimes_N\rho\, Q_\lambda^{(1)})=\det\Delta/\det\Gamma\,, \tag{3.4}$$

with Δ, Γ given by (2.10), (2.11), respectively, and $\mu_k=0$ for $k=1,2,...,n$.

Proof: The proof will be done by induction with respect to n. For $n=1$ the formula (3.4) is obviously valid. We assume the validity of (3.4) for $\dim\mathscr{H}\leqslant n$. Then for the dimension of $\mathscr{H}$ equal to $n+1$ (3.2)–(3.4) imply

$$\begin{aligned}\mathrm{tr}(&\otimes_N\rho\, Q_\lambda^{(1)})\\ &=\sum_{K\in\mathscr{K}_\lambda^{(1)}}\rho_1^{N_1^{(K)}}\cdots\rho_n^{N_n^{(K)}}\rho_{n+1}^{N_{n+1}^{(K)}}\\ &=\sum_{x_1=0}^{\lambda_1-\lambda_2}\sum_{x_2=0}^{\lambda_2-\lambda_3}\cdots\sum_{x_n=0}^{\lambda_n-\lambda_{n+1}}\Big(\sum_{K'\in\mathscr{K}_{\lambda'}^{(1)}}\rho_1^{N_1^{(K')}}\cdots\rho_n^{N_n^{(K')}}\Big)\\ &\quad\times\rho_{n+1}^{\lambda_{n+1}+x_1+\cdots+x_n}.\end{aligned} \tag{3.5}$$

Equation (3.5) is explained by the following tableau with the numbers $1,2,...,n,n+1$ distributed in a standard way. The number $n+1$ may appear only in the distinguished part of the tableau. Here $\mathscr{K}_{\lambda'}^{(1)}$ is a set of sequences $(k_1',...,k_N')$ with $k_j'=1,2,...,n$, which can be distributed in the Young frame $\lambda'=(\lambda_1-x_1,\lambda_2-x_2,...,\lambda_n-x_n)$ in the standard way



(3.6)

Using the assumption one obtains

$$\begin{aligned}\mathrm{tr}(&\otimes_N\rho\, Q_\lambda^{(1)})\\ &=\sum_{x_1=0}^{\lambda_1-\lambda_2}\cdots\sum_{x_n=0}^{\lambda_n-\lambda_{n+1}}\Big(\frac{\det\Delta'}{\det\Gamma'}\Big)\rho_{n+1}^{\lambda_{n+1}+x_1+\cdots+x_n},\end{aligned} \tag{3.7}$$

where Δ', Γ' are given by (2.10) and (2.11), respectively, with $\mu_k'=0$, $k=1,2,...,n$, and

$$\begin{aligned}\nu_k'&=(\lambda_k-x_k)+n-k\\ &=\lambda_k+(n+1)-k-(x_k+1)\\ &=\nu_k-(x_k+1).\end{aligned} \tag{3.8}$$

Using the definition of a determinant one can rewrite the rhs of (3.7) as

$$\begin{aligned}&\frac{\rho_{n+1}^{\lambda_{n+1}}}{\det\Gamma'}\sum_{p\in S_n}\epsilon_p\prod_{j=1}^n\Big\{\rho_{p(j)}^{\nu_j-1}\sum_{x_j=0}^{\lambda_j-\lambda_{j+1}}\Big(\frac{\rho_{n+1}}{\rho_{p(j)}}\Big)^{x_j}\Big\}\\ &=\Big\{\det\Gamma'\prod_{j=1}^n(\rho_j-\rho_{n+1})\Big\}^{-1}\rho_{n+1}^{\lambda_{n+1}}\\ &\quad\times\sum_{p\in S_N}\epsilon_p\prod_{j=1}^n(\rho_{p(j)}^{\nu_j}-\rho_{p(j)}^{\nu_{j+1}}\rho_{n+1}^{\nu_j-\nu_{j+1}}).\end{aligned} \tag{3.9}$$

Using now (2.12) one can transform the rhs of (3.10) into

$$(\det\Gamma)^{-1}\det\Theta\,, \tag{3.10}$$

where

$$\Theta=\begin{pmatrix}\rho_1^{\nu_1}-\rho_1^{\nu_2}\rho_{n+1}^{\nu_1-\nu_2},...,\rho_n^{\nu_1}-\rho_n^{\nu_2}\rho_{n+1}^{\nu_1-\nu_2}, & 0\\ \vdots & \vdots\\ \rho_1^{\nu_n}-\rho_1^{\nu_{n+1}}\rho_{n+1}^{\nu_n-\nu_{n+1}},...,\rho_n^{\nu_n}-\rho_n^{\nu_{n+1}}\rho_{n+1}^{\nu_n-\nu_{n+1}}, & 0\\ \rho_1^{\nu_{n+1}},...,\rho_n^{\nu_{n+1}}, & \rho_{n+1}^{\nu_{n+1}}\end{pmatrix}. \tag{3.11}$$

We multiply now the $(n+1)$th row of Θ by $\rho_{n+1}^{\nu_n-\nu_{n+1}}$ and add it to the nth row. Then we multiply the nth row by $\rho_{n+1}^{\nu_{n-1}-\nu_n}$ and add it to the $(n-1)$th row. Repeating this procedure n times we transform the matrix Θ into the matrix Δ without changing the determinant. This completes the proof of the Lemma. ■

Proof of Theorem 1: The statement (i) of Theorem 1 for the general case of possible degenerated eigenvalues of ρ may be easily obtained from the Lemma using the continuity of $\mathrm{tr}(\otimes_N\rho\, Q_\lambda^{(1)})$ with respect to $\rho_1,\rho_2,...,\rho_n$ and applying l' Hôpital's rule for the expression $\det\Delta/\det\Gamma$. Statement (ii) follows from (2.3), (2.4), and (i). ■

We present now the sketch of the proof of Theorem 2. First, we consider the case of nondegenerated eigenvalues, i.e., $p_1>p_2>\cdots>p_n$. The set of relevant Young frames $\Lambda^{(N)}$ may be identified with a subset $\Omega^{(M)}$ of the $(n-1)$-dimensional lattice $\mathbf{Z}^{n-1}$ by the following relation:

$$\begin{aligned}&\Lambda^{(N)}\ni\lambda\leftrightarrow\nu\in\Omega^{(M)}\subset\mathbf{Z}^{n-1},\\ &\nu_k=\lambda_k+n-k,\quad k=1,2,...,n-1,\\ &M=N+n(n-1)/2.\end{aligned} \tag{3.12}$$

Hence

$$\Omega^{(M)}=\Big\{\nu;\ \nu=(\nu_1,\nu_2,...,\nu_{n-1}),\ \nu_1>\nu_2>\cdots>\nu_{n-1}>M-\sum_{k=1}^{n-1}\nu_k\Big\}.$$

The probability measure $\mathscr{P}^{(N)}(\lambda)=\mathrm{tr}(\otimes_N\rho\, P_\lambda)$ on $\Lambda^{(N)}$ defines a probability measure on $\mathbf{Z}^{n-1}$, which may be written as

$$\mathscr{P}^{(M)}(\nu)=\begin{cases}W^{(M)}(\nu), & \text{for } \nu\in\Omega^{(M)},\\ 0, & \text{for } \nu\notin\Omega^{(M)},\end{cases} \tag{3.13}$$

where by Theorem 1, (2.12) and the definition of a determinant

$$W^{(M)}(\nu)=\sum_{q\in S_n}\epsilon_q G^{(M)}(\nu)W_q^{(M)}(\nu), \tag{3.14}$$

with

$$G^{(M)}(\nu)=\frac{(M-n(n-1)/2)!}{M!}\prod_{k<j=2}^{n}\frac{\nu_k-\nu_j}{p_k-p_j}, \tag{3.15}$$

$$W_q^{(M)}(\nu)=\frac{M!}{\nu_1!\nu_2!\cdots\nu_n!}p_{q(1)}^{\nu_1}p_{q(2)}^{\nu_2}\cdots p_{q(n)}^{\nu_n},$$
$$\nu_n=M-\sum_{k=1}^{n-1}\nu_k. \tag{3.16}$$

Theorem 2 is formulated in terms of the limit of distribution function for the standarized variable $\xi^{(N)}=\xi^{(N)}(\lambda)$ (2.13). The convergence of the distribution function follows from the pointwise convergence of the characteristic function

$$\Phi^{(N)}(t)=\sum_{\lambda\in\Lambda^{(N)}}e^{it\xi^{(N)}(\lambda)}\mathscr{P}^{(N)}(\lambda), \tag{3.17}$$

with

$$t=(t_1,t_2,\ldots,t_{n-1})\in\mathbb{R}^{n-1},\quad t\xi=\sum_{k=1}^{n-1}t_k\xi_k.$$

We present the idea of the proof that

$$\lim_{N\to\infty}\Phi^{(N)}(t)=\Phi_0(t),\quad \text{for all } t\in\mathbb{R}^{n-1}, \tag{3.18}$$

where $\Phi_0(t)$ is a characteristic function of the Gaussian probability distribution defined in Theorem 2 (for the nondegenerated case).

We introduce a new variable $\eta^{(M)}$ such that

$$\eta_k^{(M)}=(\nu_k-Mp_k)/(Mp_k(1-p_k))^{1/2},$$
$$k=1,2,\ldots,n-1. \tag{3.19}$$

Because $\xi^{(N)}=\eta^{(M)}+\sigma(N^{-1/2})$ then

$$\lim_{N\to\infty}\Phi^{(N)}(t)=\lim_{M\to\infty}\Psi^{(M)}(t) \tag{3.20}$$

with

$$\Psi^{(M)}(t)=\sum_{\nu\in\Omega^{(M)}}e^{it\eta^{(M)}(\nu)}\mathscr{P}^{(M)}(\nu). \tag{3.21}$$

Any term in (3.21) is dominated by the polynomial probability distribution $W_q^{(M)}(\nu)$. According to the central limit theorem (cf. Ref. 9, Theorem 6.13.1) $W_q^{(M)}(\nu)$ has a sharp maximum corresponding to $\nu^{(q,M)}\cong(Mp_{q(1)},Mp_{q(2)},\ldots,Mp_{q(n-1)})$ and the width proportional to $M^{1/2}$. Hence only for the trivial permutation $q=e$ $[e(k)=k]$ the point $\nu^{(e,M)}$ belongs to $\Omega^{(M)}$ and moreover all points $\nu^{(q,M)}$ are placed at the distances from the boundary of $\Omega^{(M)}$ proportional to M. As a consequence we obtain

$$\lim_{M\to\infty}\sum_{\nu\in\Pi^{(M)}\setminus\Omega^{(M)}}|G^{(M)}(\nu)|W_e^{(M)}(\nu)=0, \tag{3.22}$$

with

$$\Pi^{(M)}=\left\{\nu;\ \nu\in\mathbb{Z}^{n-1},\ \nu_k>0,\ \sum_{k=1}^{n-1}\nu_k<M\right\}\supset\Omega^{(M)}$$

and

$$\lim_{M\to\infty}\sum_{\nu\in\Omega^{(M)}}|G^{(M)}(\nu)|W_q^{(M)}(\nu)=0,\quad \text{for } q\in S_n,\ q\neq e. \tag{3.23}$$

It follows that

$$\lim_{M\to\infty}\Psi^{(M)}(t)=\lim_{M\to\infty}\Psi_0^{(M)}(t) \tag{3.24}$$

with

$$\Psi_0^{(M)}(t)=\sum_{\nu\in\Pi^{(M)}}e^{it\eta^{(M)}(\nu)}G^{(M)}(\nu)\left(\frac{M!}{\nu_1!\cdots\nu_n!}p_1^{\nu_1}\cdots p_n^{\nu_n}\right). \tag{3.25}$$

The function $\Psi_0^{(M)}(t)$ may be easily calculated by differentiation of the characteristic function for the polynomial distribution. From (3.25) we have

$$\Psi_0^{(M)}(t)=\tilde{G}^{(M)}(\nabla)\sum_{\nu\in\Pi^{(M)}}e^{it\eta^{(M)}}$$
$$\times\left(\frac{M!}{\nu_1!\cdots\nu_n!}p_1^{\nu_1}\cdots p_n^{\nu_n}\right), \tag{3.26}$$

where

$$\tilde{G}^{(M)}(\nabla)=\frac{(M-n(n-1)/2)!}{M!}$$
$$\times\prod_{k<j=2}^{n}(p_k-p_j)^{-1}(\nabla_k-\nabla_j),$$
$$\nabla_k=-i(Mp_k(1-p_k))^{1/2}\frac{\partial}{\partial t_k}+Mp_k, \tag{3.27}$$
$$k=1,2,\ldots,n-1,$$
$$\nabla_n=M-\sum_{k=1}^{n-1}\nabla_k.$$

Then using the explicit expressions on the characteristic function for the polynomial distribution[9] and next using the explicit expression on the limit of this function (see Ref. 9, Theorem 6.13.1) one can calculate a pointwise limit $\lim_{M\to\infty}\Psi_0^{(M)}(t)$, which is equal to $\Phi_0(t)$ [see Eq. (3.18)]. This completes the proof of Theorem 2 for the case of nondegenerated eigenvalues of ρ. Its extension to the case of degenerated eigenvalues is not difficult. The main difference is that the maximum of the relevant polynomial distribution lies on the boundary of $\Omega^{(M)}$ for $M\to\infty$. Hence the standardized variables $\{\xi_k^{(M)}\}$ corresponding to degenerated $\{p_k, k\in D_\alpha\}$ remain ordered in the limit $M\to\infty$. The formula (3.14) is also modified. After straightforward but lengthy calculations one obtains the leading term given by

$$\frac{(M-n(n-1)/2)!}{M!}(\det\Gamma)^{-1}\prod_{d=1}^{s}\bar{p}_\alpha^{-d_\alpha(d_\alpha-1)/2}$$
$$\times\prod_{1<i<j<n}(\nu_i-\nu_j)\prod_{\alpha=1}^{s}\prod_{\substack{k<l\\ k,l\in D_\alpha}}(\nu_k-\nu_l)W_e^{(M)}(\nu). \tag{3.28}$$

Then applying similar methods as before one may calculate the suitable distribution function (2.16). Strictly speaking one obtains the limit theorem first for the probability distribution symmetrized with respect to $\{\xi_k,\ k\in D_\alpha\}$ for all $\alpha = 1,2,\ldots,s$ and one recovers the restrictions on ξ_k introducing the function $\chi_\Omega(\cdot)$. ■

ACKNOWLEDGMENT

This work was supported in part by the Polish Ministry of Science and Higher Education, project CPBP 01.03.3.3.

[1]R. H. Dicke, Phys. Rev. **93**, 99 (1954).
[2]M. Gross and S. Haroche, Phys. Rep. **93**, 301 (1982), and references therein.
[3]A. Crubellier, S. Liberman, and P. Pillet, Opt. Commun. **33**, 143 (1980).
[4]A. Crubellier, S. Liberman, D. Pavolini, and P. Pillet, J. Phys. B **18**, 3911 (1985).
[5]R. Alicki, to be published in Physica A.
[6]M. Hamermesh, *Group Theory and Its Applications to Physical Problems* (Addison–Wesley, London, 1964).
[7]A. O. Barut and R. Rączka, *Theory of Group Representations and Applications* (PWN, Warsaw, 1980), 2nd ed.
[8]H. Boerner, *Representations of Groups* (Wiley, New York, 1963).
[9]M. Fisz, *Probability Theory and Mathematical Statistics* (Wiley, New York, 1963), 3rd ed.

Representation theory of superconformal quantum mechanics

A. Bohm and M. Kmiecik
Center for Particle Theory, The University of Texas at Austin, Austin, Texas 78712
L. J. Boya
Departamento de Fisica Teórica, Facultad de Ciencias, Universidad de Zaragoza, 50009 Zaragoza, Spain

(Received 8 September 1987; accepted for publication 25 November 1987)

The analogy between the representations of SU(1,1/1) and SU(2,2/1) is explained and used to suggest the latter as a spectrum supergroup of superconformal relativistic quantum mechanics for systems with rotational degrees of freedom.

I. INTRODUCTION

When "supersymmetric quantum mechanics" came into being[1] it was not immediately clear that this would lead to a straightforward extension of the idea of the spectrum generating group[2] (SGG). This emerged only when a fusion of conformal quantum mechanics[3] and supersymmetry was attempted.[4] This led to the introduction of a pair of spinor operators S and $S^\dagger$, which—in the same way as the original spinor operators Q and $Q^\dagger$ were the square roots of the Hamiltonian—are square roots of the conformal generator $\mathfrak{K}$,

$$\tfrac{1}{2}\{Q,Q^\dagger\} = \mathfrak{H}; \quad \tfrac{1}{2}\{S,S^\dagger\} = \mathfrak{K}. \tag{1.1}$$

Here, $\mathfrak{H}$, $\mathfrak{K}$, and their commutator $\mathscr{D}$ are the generators of SO(2,1) fulfilling the commutation relations (CR)

$$[\mathfrak{H},\mathscr{D}] = i\mathfrak{H}, \quad [\mathfrak{K},\mathscr{D}] = -i\mathfrak{K}, \quad [\mathfrak{H},\mathfrak{K}] = 2i\mathscr{D}. \tag{1.2}$$

SO(2,1) was the first example of a group that describes the whole spectrum of a physical system[2] and was originally discovered as the SGG for the collective vibrations of a nucleus. Later it became clear that SO(2,1) did not only apply to the harmonic oscillator but was a more general feature of the radial motion.[5]

The remaining, nonvanishing anticommutators of the spinor operators

$$\tfrac{1}{2}\{Q,S^\dagger\} = i\mathscr{D} - \mathscr{Y}, \quad \tfrac{1}{2}\{Q^\dagger,S\} = -i\mathscr{D} - \mathscr{Y}, \tag{1.3}$$

introduce the generator $\mathscr{Y}$ ($= B/2 - f/2$ in the notation of Ref. 4) of a U(1) subgroup which commutes with SO(2,1). These four fermionic and four bosonic operators generate the supergroup Osp(2,2) $\simeq$ SU(1,1/1); all its defining relations are given below in Eq. (3.3).

Nonrelativistic supersymmetric quantum mechanics has usually been formulated in terms of this Osp(2,2) or its subsupergroup Osp(1,2). Osp(2,2) $\supset$ Osp(1,2) has been used as the spectrum generating supergroup (also called a "spectrum supersymmetry") for the one-dimensional oscillator with spin,[6] for the fermion-monopole system,[7] for the noncanonical two-particle oscillator,[8] and for other nonrelativistic systems. In the same way as the bosonic part SO(2,1) is not specific for one particular interaction but a general feature of the radial motion,[5] Osp(2,2) can also be realized in many different ways in terms of the radial distance and momentum operators and the Pauli matrices.

We shall use the superconformal quantum mechanics[4] based on

$$\begin{array}{l} SU(1,1/1)_{Q,S,\mathfrak{H},\mathfrak{K},\mathscr{D},\mathscr{Y}} \supset Osp(1,2) \supset SO(2,1)_{\mathfrak{H},\mathfrak{K},\mathscr{D}} \supset SO(2)_{(1/2)(\mathfrak{H}+\mathfrak{K})} \\ \cup \\ U(1)_{\mathscr{Y}} \times SO(2,1)_{\mathfrak{H},\mathfrak{K},\mathscr{D}} \end{array} \tag{1.4}$$

as our point of departure. Our ultimate goal is a relativistic supersymmetric quantum mechanics. For the latter we suggest SU(2,2/1) in place of SU(1,1/1). But SU(2,2/1) in a different physical interpretation than ours may also be useful for nonrelativistic systems with rotational degrees of freedom[9,10] like, e.g., the Coulomb and dyon systems which use SO(4,2) as spectrum symmetry. The analogy between SU(1,1/1) and SU(2,2/1) becomes already apparent when one writes their defining relations next to each other as done in Eq. (3.3).

The generators of a group with physical interpretation are observables with a definite physical meaning. Therefore, even if two groups are isomorphic they may be physically different. The statement that SU(1,1/1) is a subsupergroup of SU(2,2/1) is therefore of little help if one does not say which SU(1,1/1) is meant. To characterize the physical meaning of a group (or supergroup) we shall often write the generators of the group as subscripts on the symbol for the group as, e.g., done for the subgroup chain (1.4).

We will deal here exclusively with star representations, not grade-star representations,[11] of superalgebras because of the general belief that algebras of observables in quantum mechanics are star algebras.[12] This requires infinite dimensional representations. For Lie algebras and Lie groups the connection between the representations of the algebra and the unitary representations of the group is well known. It is remarkable that in physics—also for spectrum generating algebras, which are not connected with transformations of a symmetry group—only those representations of Lie algebras occur that integrate to representations of the group. For superalgebras the notion of integrability has been defined recently.[13] We will only discuss representations of superalge-

 0022-2488/88/051163-08$02.50

bras that integrate to representations of the supergroup, in the sense of Ref. 13. For these, every irreducible representation (irrep) of a superalgebra is a finite direct sum of its (noncompact) even subalgebra. The infiniteness of the representation space is not worse than for the noncompact Lie subgroup and the irrep space can be put together from a finite number (two in the case we treat here) of irrep spaces of the even subgroup.

In Sec. II we review the properties of SU(1,1/1) $\supset$Osp(1,2) and describe the class of representations that are important for supersymmetric quantum mechanics. In Sec. III we explain the transition from SU(1,1/1) to SU(2,2/1). In Sec. IV the so-called "massless" positive energy representations of SU(2,2/1) are described and the parity operator is defined. In Sec. V, the coupling of these SU(2,2/1) representations to the Poincaré group is described.

II. INFINITE DIMENSIONAL NONTYPICAL IRREPS OF SU(1,1/1)

The irreps of Osp(1,2) $\subset$SU(1,1/1) are obtained as the direct sum of two irreps of SO(2,1) from the discrete series $D_+(q)$,[14]

$$D^{\mathrm{Osp}(1,2)}(q_0) \underset{\mathrm{SO}(2,1)}{\rightleftharpoons} D_+(q_0) \oplus D_+(q_0 + \tfrac{1}{2}). \tag{2.1}$$

They are characterized by one number q_0 which can take the following values:

$$q_0 = \tfrac{1}{2}, 1, \tfrac{3}{2}, 2, \ldots . \tag{2.2}$$

The basis vectors in the irrep space $\mathfrak{H}^{\mathrm{Osp}(1,2)}$ are denoted by

$$|q_0;q,\mu\rangle, \quad q = q_0, q_0 + \tfrac{1}{2}, \quad \mu = q, q+1, q+2, \ldots . \tag{2.3}$$

The label μ is the eigenvalue of Γ_0, the compact generator of SO(2,1) given by

$$\Gamma_0 = \tfrac{1}{2}(\mathfrak{H} + \mathfrak{K}). \tag{2.4}$$

Here $q(q-1)$ is the eigenvalue of $\mathscr{C}(\mathrm{SO}(2,1))$ where $\mathscr{C}(\mathrm{SO}(2,1)) = \Gamma_0^2 - \tfrac{1}{2}\{B_+, B_-\}$ is the Casimir operator of SO(2,1) and

$$B_\pm = -\tfrac{1}{2}(\mathfrak{H} - \mathfrak{K} \mp 2i\mathscr{D}) \tag{2.5}$$

are the bosonic excitation operators,

$$B_\pm |q_0;q,\mu\rangle = \sqrt{\mu(\mu \pm 1) - q(q-1)}\,|q_0,q,\mu \pm 1\rangle. \tag{2.6}$$

Osp(1,2) has the bosonic generators $B_\pm$, Γ_0, and the fermionic generators

$$F_\pm = (1/\sqrt{2})(i/2)(S^\dagger - S \mp (Q + Q^\dagger)). \tag{2.7}$$

The fermionic excitation operators $F_\pm$ change the eigenvalues μ of Γ_0 by one half unit and also change the value of q,

$$\begin{aligned} F_\pm |q_0;q = q_0,\mu\rangle &= (1/\sqrt{2})\sqrt{\mu \pm q_0}\,|q_0;q \\ &= q_0 + \tfrac{1}{2}, \mu \pm \tfrac{1}{2}\rangle, \\ F_\pm |q_0;q = q_0 + \tfrac{1}{2},\mu\rangle &= (1/\sqrt{2})\sqrt{\mu \mp (q_0 - \tfrac{1}{2})}\,|q_0;q \\ &= q_0, \mu \pm \tfrac{1}{2}\rangle. \end{aligned} \tag{2.8}$$

With Eqs. (2.6) [for SO(2,1)] and (2.8) the representation $D(q_0)$ of Osp(1,2) is completely determined. To illustrate the properties of the irreps of supergroups one uses, as in the case of noncompact groups, the weight diagram. The weight diagram displays which irreps of the (maximal) compact subgroup, in our case $\mathrm{SO}(2)_{\Gamma_0}$, occur in an irrep of the supergroup. Figure 1 shows the weight diagram of $D(q_0)$ of Osp(1,2). Each dot represents a value of μ that occurs in $D(q_0)$. The reduction property (2.1) is displayed by the two columns of dots, each column representing the weight diagram of the irrep $D_+(q)$ of SO(2,1). The action of the generators (2.6) and (2.8) is also depicted in the weight diagram.

The representations of $\mathrm{Osp}(2,2) \simeq \mathrm{SU}(1,1/1)_{S,Q,S^\dagger,Q^\dagger,B_\pm,\Gamma_0,\mathscr{Y}}$ that we are interested in here are the nontypical representations[15] which remain irreducible when restricted to the subsupergroup $\mathrm{Osp}(1,2)_{F_\pm B_\pm,\Gamma_0}$. For each value of q_0 there are two irreps of Osp(2,2) which we denote by $D_S(\pm q_0)$. We will consider only $D_S(+q_0)$; their reduction with respect to the subgroup chain

$$\begin{aligned} &\mathrm{SU}(1,1/1) \supset \mathrm{SO}(2,1)_{B_\pm,\Gamma_0} \\ &\quad \times \mathrm{U}(1)_{\mathscr{Y}} \supset \mathrm{SO}(2)_{\Gamma_0} \times \mathrm{U}(1)_{\mathscr{Y}} \end{aligned} \tag{2.9}$$

is given by

$$\begin{aligned} D_S(q_0) &\rightleftharpoons D_+(q = q_0) \times (y = q_0) \\ &\quad \oplus D_+(q = q_0 + \tfrac{1}{2}) \times (y = q_0 + \tfrac{1}{2}), \end{aligned} \tag{2.10}$$

where (y) denotes the one-dimensional representation of $\mathrm{U}(1)_{\mathscr{Y}}$. From (2.10) one sees that for these nontypical irreps y is not an independent quantum number but is already fixed by the value of q which characterizes the irrep $D_+(q)$ of the SO(2,1) subgroup. With this, the weight diagram in Fig. 1 becomes identical with Fig. 2 of Ref. 4 (where $y = B/2 - f/2$).

In the physical applications, the dots of the mathematical weight diagram represent physical states. For example, if SO(2,1) is interpreted as the spectrum generating group of the oscillator[2,5] then μ becomes—except for an additive constant—the vibrational quantum number and each column of dots represents the equidistant energy levels of a harmonic oscillator. The weight diagram of Fig. 1 then becomes the energy diagram of the radial part of the harmonic oscillator with spin.[6] The same representation can be used for the fermion-monopole system.[7]

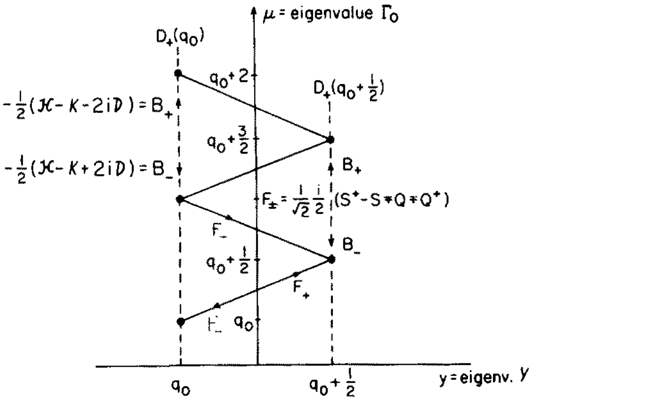


FIG. 1. Weight diagram of Osp(1,2) which is also the weight diagram of the irrep $D_S(q_0)$ of Osp(2,2) or Su(1,1/1) showing the actions of the generators.

III. FROM SU(1,1/1) TO SU(2,2/1)

A vibrator in three dimensions has in addition to the SO(2,1) dynamical group of the radial motion an $SO(3)_{S_{ij}}$ symmetry group ($i,j = 1,2,3$, $S_{ij} = \epsilon_{ijk}S_k$) describing the rotational degrees of freedom. The SO(2,1) and SU(1,1/1) must therefore be enlarged. The minimal choice for the bosonic part, which would contain SO(2,1) as the spectrum generating group and $SO(3)_{S_{ij}}$ as a symmetry group, is

$$SO(4,2) \supset SO(2,1)_{\mathfrak{P}_0\mathfrak{K}_0D} \times SO(3)_{S_{ij}}. \tag{3.1}$$

The immediate minimal choice for the superalgebra is then SU(2,2/1) because of the following similarity:

$$\begin{array}{lll} SU(2,2/1) \supset SO(4,2) & \times U(1)_{\mathscr{Y}} \supset SO(3)_{S_{ij}} \times SO(2)_{\Gamma_0} \times U(1) \\ SU(1,1/1) \supset SO(2,1)_{\mathfrak{H}\mathscr{K}\mathscr{D}} & \times U(1)_{\mathscr{Y}} \supset & SO(2)_{\Gamma_0} \times U(1). \end{array} \tag{3.2}$$

The analogy between the enlarged and the original subgroup chains becomes even more obvious when one juxtaposes their defining relations as done in Eqs. (3.3),[16]

$$\begin{array}{lll}
 & SU(2,2/1) & SU(1,1/1) \\
a & [S_{\mu\nu},S_{\rho\sigma}] = i(\eta_{\nu\rho}S_{\mu\sigma} + \eta_{\mu\sigma}S_{\nu\rho} - \eta_{\mu\rho}S_{\nu\sigma} - \eta_{\nu\sigma}S_{\mu\rho}), & \\
b & [\mathfrak{P}_\mu,S_{\rho\sigma}] = i(\eta_{\mu\rho}\mathfrak{P}_\sigma - \eta_{\mu\sigma}\mathfrak{P}_\rho), & \\
c & [\mathfrak{K}_\mu,S_{\rho\sigma}] = i(\eta_{\mu\rho}\mathfrak{K}_\sigma - \eta_{\mu\sigma}\mathfrak{K}_\rho), & \\
d & [D,S_{\mu\nu}] = 0, \quad [\mathfrak{P}_\mu,\mathfrak{P}_\nu] = 0, \quad [\mathfrak{K}_\mu,\mathfrak{K}_\nu] = 0, & \\
e & [\mathfrak{P}_\mu,D] = i\mathfrak{P}_\mu, & [\mathfrak{H},\mathscr{D}] = i\mathfrak{H}, \\
f & [\mathfrak{K}_\mu,D] = -i\mathfrak{K}_\mu, & [\mathfrak{K},\mathscr{D}] = -i\mathfrak{K}, \\
g & [\mathfrak{P}_\mu,\mathfrak{K}_\nu] = 2i(\eta_{\mu\nu}D - S_{\mu\nu}), & [\mathfrak{H},\mathfrak{K}] = 2i\mathscr{D}, \\
h & [Q,S_{\mu\nu}] = \tfrac{1}{2}\sigma_{\mu\nu}Q, \quad [S,S_{\mu\nu}] = \tfrac{1}{2}\sigma_{\mu\nu}S, & \\
i & [Q,D] = \tfrac{1}{2}iQ, \quad [S,D] = -\tfrac{1}{2}iS, & [\widetilde{Q}_\alpha,\mathscr{D}] = (i/2)\widetilde{Q}_\alpha, \quad [\widetilde{S}_\alpha,\mathscr{D}] = -(i/2)\widetilde{S}_\alpha, \\
j & [Q,\mathfrak{P}_\mu] = 0, \quad [S,\mathfrak{P}_\mu] = \gamma_\mu Q, & [\widetilde{Q}_\alpha,\mathfrak{K}] = 0, \quad [\widetilde{S}_\alpha,\mathfrak{K}] = +i\widetilde{Q}_\alpha, \\
k & [Q,\mathfrak{K}_\mu] = \gamma_\mu S, \quad [S,\mathfrak{K}_\mu] = 0, & [\widetilde{Q}_\alpha,\mathfrak{K}] = -i\widetilde{S}_\alpha, \quad [\widetilde{S}_\alpha,\mathfrak{K}] = 0, \\
l & [Q,Y] = -i\tfrac{3}{2}\gamma_5 Q, \quad [S,Y] = i\tfrac{3}{2}\gamma_5 S, & [\widetilde{Q}_\alpha,\mathscr{Y}] = -(i/2)\epsilon_{\alpha\beta}\widetilde{Q}_\beta, \quad [\widetilde{S}_\alpha,\mathscr{Y}] = (i/2)\epsilon_{\alpha\beta}\widetilde{S}_\beta, \\
m & [Y,\mathfrak{P}_\mu] = [Y,\mathfrak{K}_\mu] = [Y,D] = [Y,S_{\mu\nu}] = 0, & [\mathscr{Y},\mathfrak{H}] = [\mathscr{Y},\mathfrak{K}] = [\mathscr{Y},\mathscr{D}] = 0, \\
n & \{Q,\overline{Q}\} = 2\gamma^\mu\mathfrak{P}_\mu, \quad \{S,\overline{S}\} = 2\gamma^\mu\mathfrak{K}_\mu, & \{\widetilde{Q}_\alpha,\widetilde{Q}_\beta\} = 2\delta_{\alpha\beta}\mathfrak{H}, \quad \{\widetilde{S}_\alpha,\widetilde{S}_\beta\} = 2\delta_{\alpha\beta}\mathfrak{K}, \\
o & \{S,\overline{Q}\} = 2iD + \sigma^{\mu\nu}S_{\mu\nu} - 2i\gamma_5 Y, & \{\widetilde{S}_\alpha,\widetilde{Q}_\beta\} = -2\delta_{\alpha\beta}\mathscr{D} - 2\epsilon_{\alpha\beta}\mathscr{Y}.
\end{array}$$

with $\eta_{ab} = (+1,-1,-1,-1,-1,+1)$ for $a = b = 0,1,2,3,5,6$.

The operators Q and S in the defining CR (3.3) of SU(2,2/1) are two Majorana spinors,

$$Q = \begin{pmatrix} Q_1 \\ Q_2 \\ Q_2^\dagger \\ -Q_1^\dagger \end{pmatrix} = \begin{pmatrix} Q_\alpha \\ \overline{Q}^{\dot\alpha} \end{pmatrix}, \quad S = \begin{pmatrix} S_1 \\ S_2 \\ S_2^\dagger \\ -S_1^\dagger \end{pmatrix} = \begin{pmatrix} S_\alpha \\ \overline{S}^{\dot\alpha} \end{pmatrix}, \tag{3.4}$$

and $\widetilde{Q}_\alpha$ and $\widetilde{S}_\alpha$ for SU(1,1/1) are linear combinations of the Q, $Q^\dagger$, S, $S^\dagger$ in Eq. (1.1). These linear combinations are

$$\begin{array}{l} \widetilde{Q}_1 = (1/\sqrt{2})(Q + Q^\dagger); \quad \widetilde{Q}_2 = -(i/\sqrt{2})(Q - Q^\dagger); \\ \widetilde{S}_1 = (i/\sqrt{2})(S^\dagger - S); \quad \widetilde{S}_2 = -(1/\sqrt{2})(S^\dagger + S). \end{array} \tag{3.5}$$

The relations (3a)–(3g) are the defining relations of the subgroup $SO(4,2) \simeq SU(2,2)$. There is another standard notation for the basis of the Lie algebra of SO(4,2) that is more useful for our purposes. One defines

$$\begin{array}{l} S_{\mu5} = \tfrac{1}{2}(\mathfrak{P}_\mu - \mathfrak{K}_\mu), \quad S_{\mu6} \equiv \Gamma_\mu = \tfrac{1}{2}(\mathfrak{P}_\mu + \mathfrak{K}_\mu), \\ S_{\mu\nu} \quad (\mu,\nu = 0,1,2,3). \end{array} \tag{3.6}$$

The S_{ab} ($a,b = 0,1,2,3,5,6$) fulfill as a consequence of (3.3) the commutation relations

$$[S_{ab},S_{cd}] = -i(\eta_{ac}S_{bd} + \eta_{bd}S_{ac} - \eta_{ad}S_{bc} - \eta_{bc}S_{ad}) \tag{3.7}$$

With the superalgebra SU(2,2/1) chosen, the first attempt to generalize the subgroup chain (3.1) to a supersubgroup chain would be to extend $SO(2,1)_{\mathfrak{P}_0\mathfrak{K}_0D}$ to an SU(1,1/1) by adjoining the $\widetilde{Q}_\alpha$, $\widetilde{Q}_\alpha^\dagger$, $\widetilde{S}_\alpha$, $\widetilde{S}_\alpha^\dagger$ and to extend $SO(4,2)_{\mathfrak{P}_\mu\mathfrak{K}_\mu S_{\mu\nu}D}$ to an SU(2,2/1) by adjoining the two Majorana spinors Q, S,

$$\begin{array}{lll} SO(4,2)_{\mathfrak{P}_\mu\mathfrak{K}_\mu S_{\mu\nu}D} & \supset & SO(2,1)_{\mathfrak{P}_0\mathfrak{K}_0D} \times SO(3)_{S_{ij}} \\ \big\downarrow \text{ adjoin } Q,S & & \big\downarrow \text{ adjoin } \widetilde{Q}_\alpha,\widetilde{S}_\alpha \\ SU(2,2/1) & & SU(1,1/1)_{\mathfrak{P}_0,\mathfrak{K}_0,D,\widetilde{Q}_\alpha,\widetilde{S}_\alpha} \times SO(3)_{S_{ij}}. \end{array} \tag{3.8}$$

Unfortunately this *cannot* be done in such a way that $SU(1,1/1)_{\mathfrak{P}_0\mathfrak{K}_0D\widetilde{Q}\widetilde{S}}$ is a subsuperalgebra of $SU(2,2/1)_{\mathfrak{P}_\mu\mathfrak{K}_\mu S_{\mu\nu}D,Q,S}$. In other words one *cannot* find a linear combination of the four spinor operators $\widetilde{Q}_\alpha$, $\widetilde{S}_\alpha$, $\alpha = 1,2$, in terms of the eight components Q_α, $\overline{Q}^{\dot\alpha}$, S_α, $\overline{S}^{\dot\alpha}$ ($\alpha,\dot\alpha = 1,2$) of the Majorana spinor such that $\widetilde{Q}_\alpha$, $\widetilde{S}_\alpha$ together with $\mathfrak{P}_0$, $\mathfrak{K}_0$, and D fulfill the commutator–anticommutator relation of $Osp(2,1) \subset SU(1,1/1)$. SU(1,1/1) is a subsupergroup of SU(2,2/1) but the generators of this $SU(1,1/1)_{\mathfrak{H}\mathfrak{K}\mathscr{D}\widetilde{Q}_\alpha\widetilde{S}_\alpha}$ subsupergroup of $SU(2,2/1)_{\mathfrak{P}_\mu\mathfrak{K}_\mu S_{\mu\nu}DQS}$ are not $\mathfrak{P}_0$, $\mathfrak{K}_0$, D, and linear combinations of Q_α, $\overline{Q}^{\dot\alpha}$, S_α, $\overline{S}^{\dot\alpha}$.

One can find such an SU(1,1/1) subsupergroup, which is, e.g., generated by

$$\mathfrak{H} = \tfrac{1}{2}(\mathfrak{P}_0 + \mathfrak{P}_3), \quad \mathfrak{K} = \tfrac{1}{2}(\mathfrak{K}_0 - \mathfrak{K}_3),$$
$$\mathscr{D} = \tfrac{1}{2}(D + S_{03}); \quad \mathscr{Y} = \tfrac{1}{2}(S_{12} + Y), \tag{3.9a}$$

and

$$\widetilde{Q}_1 = \tfrac{1}{2}(Q_1 + Q_1^\dagger), \quad \widetilde{Q}_2 = -(i/2)(Q_1 - Q_2^\dagger),$$
$$\widetilde{S}_1 = (i/2)(S_2^\dagger - S_2), \quad \widetilde{S}_2 = \tfrac{1}{2}(S_2^\dagger + S_2), \tag{3.9b}$$

where the Q_1, $Q_1^\dagger$, and S_2, $S_2^\dagger$ are the components of the Majorana spinor (3.4). This SU(1,1/1) gives, however, other troubles.

When one extends the dynamical group chain $SO(2,1) \supset SO(2)_{\Gamma_0}$ to include rotations, the physical states, labeled by μ and y (or q) and represented by the • in Fig. 1, will have to be labeled in addition by the angular momentum j where $j(j+1) =$ eigenvalue of $(\tfrac{1}{2}S_{ij}S_{ij})$. This is not possible if one chooses the $SU(1,1/1)_{\mathfrak{H}\mathscr{K}\mathscr{D}\widetilde{Q}_a\widetilde{S}_a}$ with the generators (3.9) because this SU(1,1/1) does not commute with $SO(3)_{S_{ij}}$. Therefore one must give up the requirement that the SU(2,2/1) have an $SU(1,1/1)\times SO(3)_{S_{ij}}$ as a subsupergroup.

But this requirement is not necessary if one wants to extend the spectrum described by the diagram in Fig. 1 to a spectrum that includes angular momentum. All that one needs is a representation of SU(2,2/1) that reduces with respect to the subgroup chain

$$\begin{aligned} SU(2,2/1) &\supset SO(4,2)_{S_{ab}} \times U(1)_{\mathscr{Y}} \\ &\supset SO(2,1)_{S_{06},S_{05},S_{56}} \times SO(3)_{S_{ij}} \\ &\supset SO(2)_{\Gamma_0} \times SO(3)_{S_{ij}} \end{aligned} \tag{3.10}$$

such that it contains the representations $D_+(q = q_0)$ and $D_+(q = q_0 + \tfrac{1}{2})$ of $SO(2,1)_{S_{06},S_{05},S_{56}} = SO(2,1)_{\mathfrak{P}_0\mathfrak{K}_0 D}$ depicted by the two towers (columns of dots) in Fig. 1. Each tower is to be labeled in addition by a definite value of angular momentum j coming from the $SO(3)_{S_{ij}}$ and preferably there are not to be any other labels of the states besides μ (radial quantum number), j (angular momentum quantum number), and (its component) j_3. Representations of SU(2,2/1) with these properties do indeed exist and are contained in the list of representations classified in Ref. 17. Their specific properties that we need follow immediately from their reduction with respect to the bosonic subgroup $SO(4,2)\times U(1)$ and we present them in the following section.

IV. THE SO-CALLED "MASSLESS" "POSITIVE ENERGY" REPRESENTATIONS OF SU(2,2/1)

These representations get their name from the fact that in their representation spaces the operators have the following properties:

$$\mathfrak{P}_\mu \mathfrak{P}^\mu = 0, \quad \text{spectrum } \mathfrak{P}_0 > 0 \Rightarrow \text{spectrum } \Gamma_0 > 0. \tag{4.1}$$

In place of (1.4) we have now the subsupergroup chain

$$\begin{aligned} SU(2,2/1) &\supset Osp(1,4)_{S_{\mu\nu}\Gamma_\mu Q_I} \supset SO(2,3)_{S_{\mu\nu}\Gamma_\mu} \\ &\supset SO(2)_{\Gamma_0} \times SO(3)_{S_{ij}}, \end{aligned} \tag{4.2}$$

where the Majorana spinor operator Q_I is given in terms of the operators (3.4) by

$$Q_I = \tfrac{1}{2}(Q + S). \tag{4.3}$$

All representations of SU(2,2/1), Osp(1,4), SU(2,2), or SO(3,2) with spectrum $\Gamma_0 > 0$ (or < 0) are characterized by their lowest weight.[17,18] The weights of a representation are the sets of numbers which characterize the irreps of the maximal compact subgroup K. The weight of $SO(3,2) \subset Osp(1,4)$ is the pair of numbers (μ, j) where μ characterizes the $SO(2)_{\Gamma_0}$ irrep and j the irrep of SO(3). The irreps of SO(3,2) and Osp(1,4) are thus denoted by $D(\mu^0, j^0)$ and $D_S(\mu^0, j^0)$, respectively, where (μ^0, j^0) is the "lowest" of the pairs (μ, j). The weight of SU(2,2) is the triplet of numbers $(\mu; j^{(1)}, j^{(2)})$, where μ is as above and $j^{(1)}$, $j^{(2)}$ characterize the irreps of SO(4). [The maximal compact subgroup of SU(2,2) is $SO(2)_{\Gamma_0} \times SO(4)_{S_5,S_{ij}}$.] The irreps of SU(2,2) are thus denoted by $D(\mu^0; j^{(1)0} j^{(2)0})$, where $\mu^0 j^{(1)0} j^{(2)0}$ is the lowest weight.[18] The representations that fulfill (4.1) are all of the kind $D(s+1; s, 0)$ or $D(s+1; 0, s)$, where s is an integer or half-integer $s = 0$ or $\tfrac{1}{2}$ or 1 or $\tfrac{3}{2}$ or $\cdots$.

The "massless" "positive energy" representations of the supergroups SU(2,2/1) and Osp(1,4) have been obtained in a series of papers.[17] They are characterized by one number s_0 which is an integer or half-integer. There are two classes of irreps and two (inequivalent) irreps for each value s_0. We first restrict ourselves to the irreps denoted by $D_S(s_0 + 1; s_0, 0; s_0 + 1)$. There reduction to the bosonic subgroup $SU(2,2)_{S_{ab}} \times U(1)_Y$ is

$$\begin{aligned} &D_S(s_0 + 1; s_0, 0; s_0 + 1) \\ &\xrightarrow[SO(4,2)\times U(1)]{} D(s_0 + 1; s_0, 0) \times [s_0 + 1] \\ &\qquad \oplus D(s_0 + \tfrac{3}{2}; s_0 + \tfrac{1}{2}, 0) \times [s_0 - \tfrac{1}{2}], \end{aligned} \tag{4.4}$$

where $[s_0 + 1]$, $[s_0 - \tfrac{1}{2}]$ denotes the $U(1)_Y$ character. Like in the reduction of the nontypical irreps of SU(1,1/1) with respect to Osp(1,2), these irreps of SU(2,2/1) also remain irreducible when restricted to the subsupergroup Osp(1,4). The Osp(1,4) representations are therefore also characterized by the same number s_0 and are denoted by $D_S(s_0 + 1, s_0)$. Their reduction with respect to the bosonic subgroup SO(3,2) is given by

$$D_S(s_0 + 1, s_0) \xrightarrow[SO(3,2)]{} D(s_0 + 1, s_0) \oplus D(s_0 + \tfrac{3}{2}, s_0 + \tfrac{1}{2}). \tag{4.5}$$

In the special class of representations $D(s+1; s, 0)$—and therefore also in $D_S(s_0 + 1; s_0, 0; s_0 + 1)$—the following relation holds:

$$E_{ab} \equiv -\{S_{ac}, S^c{}_b\} - \eta_{ab}\tfrac{1}{3}S_{cd}S^{cd} = 0. \tag{4.6}$$

[E_{ab} generate an ideal of the enveloping algebra of SO(4,2) which is the kernel that defines the massless representations; all properties of this representation of SO(4,2) can be derived from the relation (4.6).[19]] A consequence of (4.6) is the following relation between the Casimir operators of the subgroups SO(2,1) and SO(3):

$$\begin{aligned} \mathscr{C}(SO(2,1)) &\equiv -S_{05}^2 - S_{56}^2 + \Gamma_0^2 \\ &= \tfrac{1}{2}S_{ij}S^{ij} \equiv \mathscr{C}(SO(3)). \end{aligned} \tag{4.7}$$

For the eigenvalues of these two Casimir operators this relation means

$$q(q-1)=j(j+1) \quad \text{or} \quad q=j+1. \tag{4.7'}$$

Thus in $D_S(s_0+1;s_0,0;s_0+1)$ the representations of the subgroups SO(2,1) and SO(3) are related to each other. Each tower of radial excitations $D_+(q)$ has a definite value of angular momentum j, namely $j=q-1$.

According to (4.5) and (4.4), the weight diagram of the supergroup is a combination of two weight diagrams for its bosonic subgroup. Equation (4.4) is the analog of (2.10) and Eq. (4.5) is the analog of (2.1).

Figure 2 shows the weight diagram for the special case $s_0=\frac{1}{2}$,

$$D_S(\tfrac{3}{2};\tfrac{1}{2},0;\tfrac{3}{2}) \underset{\mathrm{SO}(4,2)}{\Longrightarrow} D(\tfrac{3}{2};\tfrac{1}{2},0) \oplus D(2;1,0). \tag{4.8}$$

The diagram in Fig. 2 displays the reduction of $D_S(\frac{3}{2};\frac{1}{2},0;\frac{3}{2})$ with respect to the subgroup chain (4.2). Here μ is plotted versus j and each dot stands for an irrep space of $\mathrm{SO}(2)_{\Gamma_0} \times \mathrm{SO}(3)_{S_{ij}}$ representing states with definite angular momentum j and radial excitation μ. Each column of dots with a given value j stands for an irrep $D_+(q)$ of $\mathrm{SO}(2,1)_{\Gamma_0,S_{05},S_{56}}$ with $q=j+1$ representing all radial excitations with the same angular momentum j. The bosonic generators of SO(2,1), $B_\pm = S_{05} \pm iS_{56} = -\frac{1}{2}(\mathfrak{P}_0 - \mathfrak{K}_0 \mp 2iD)$, again transform along the vertical changing the value of μ (radial excitation number) by one unit without changing the value of j.

The fermionic generators that change μ by one-half unit are now given by the components of the Majorana spinor Q_I of (4.3). They also change the value of j by one-half unit. The two columns leftmost in the weight diagram (or any two adjacent columns) of Fig. 2 are thus identical to the weight diagram of Fig. 1 with $q_0=s_0+1=\frac{3}{2}$. The bosonic generators $B_\pm$ transform between the states like the corresponding generators in Fig. 1. In addition there are now the bosonic generators Γ_i and S_{0i} which transform along the diagonal and change μ and j by one unit (except for $s=0$ they also

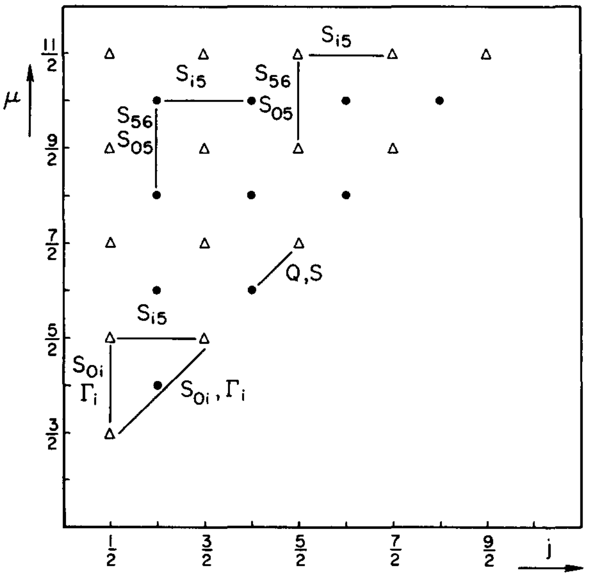


FIG. 2. Weight diagram of the irrep $D_S(s+1;s,0;s+1)$ of SU(2,2/1) for the special case $s=\frac{1}{2}$. The actions of the generators and the reduction

$$D_S(\tfrac{3}{2};\tfrac{1}{2},0;\tfrac{3}{2}) \underset{\mathrm{SO}(4,2)}{\Rightarrow} D(\tfrac{3}{2};\tfrac{1}{2},0) \oplus D(2,1,0)$$

are also shown.

transform along the vertical). And there are the bosonic generators S_{i5} which transform along the horizontal changing j by one unit and keeping μ fixed. The S_{i5} together with the S_{ij} generate the $\mathrm{SO}(4)_{S_{ij},S_{i5}}$. The finite number of dots along the horizontal $\mu=$ const stand for an irrep of SO(4) and display its reduction with respect to $\mathrm{SO}(3)_{S_{ij}}$. The weight diagram of Fig. 2 is thus an extension of the weight diagram of Fig. 1 from two towers of radial excitations to an infinite number of towers of radial excitations, one for each half-integer and integer value of j.

We will now discuss the basis vectors of the representation space of $D_S(s_0+1;s_0,0;s_0+1)$. Each dot in Fig. 2 corresponds to an irrep space of $\mathrm{SO}(2)_{\Gamma_0} \times \mathrm{SO}(3)_{S_{ij}}$ which we will denote by $\mathfrak{H}^\mu(j)$ where the radial quantum number μ and $j=q-1$ are fixed and correspond to the position of the dot in the weight diagram of Fig. 2. The basis vectors, which consist of eigenvectors of Γ_0, $\mathbf{S}^2$, S_{12}, are denoted by

$$|\mu,j,j_3;+s_0\rangle, \quad -j \leqslant j_3 \leqslant j. \tag{4.9}$$

The direct sum over all $\mathfrak{H}^\mu(j)$ with (μ,j) being a dot at integer values in Fig. 2 is the bosonic space of states

$$\mathfrak{H}_+^{+s} = \sum_{\substack{\mu=s+1,s+2,\ldots\\ j=s,s+1,\ldots,\mu-1}}^{\infty} \oplus \mathfrak{H}^\mu(j), \quad \text{where } s=\text{integer}. \tag{4.10}$$

The direct sum of all $\mathfrak{H}^\mu(j)$, where (μ,j) is the triangle at half-integer values in Fig. 2, is the fermionic space,

$$\mathfrak{H}_-^{+s} = \sum_{\substack{\mu=s+1,s+2,\ldots\\ j=s,s+1,\ldots,\mu-1}}^{\infty} \oplus \mathfrak{H}^\mu(j), \quad \text{where } s=\text{half-integer}. \tag{4.11}$$

These are irreducible representation spaces of SO(4,2) characterized by the value s. The space

$$\mathfrak{H}^{+s_0} = \mathfrak{H}^{s=s_0} \oplus \mathfrak{H}^{s=s_0+1/2} \tag{4.12}$$

is the irrep space of SU(2,2/1). (If $s_0=\frac{1}{2},\frac{3}{2},\ldots$, $\mathfrak{H}^{s_0}=\mathfrak{H}_-$ and $\mathfrak{H}^{s_0+1/2}=\mathfrak{H}_+$, and if $s_0=1,2,3,\ldots$, $\mathfrak{H}^{s_0}=\mathfrak{H}_+$ and $\mathfrak{H}^{s_0+1/2}=\mathfrak{H}_-$.) The spinor operators Q_I and $Q' \equiv \frac{1}{2}\gamma_5(Q-S)$ transform between $\mathfrak{H}_+$ and $\mathfrak{H}_-$, the bosonic operators S_{ab} transform within an $\mathfrak{H}_+$ or $\mathfrak{H}_-$.

The vectors (4.9) are in general not eigenvectors of the parity operator U_P [and the $\mathfrak{H}^\mu(j)$ are in general not eigenspaces of U_P]. In order to obtain parity eigenstates we have to discuss also the other class of massless, positive energy representations of SU(2,2/1). These representations are denoted by $D_S(s_0+1;0,s_0;-s_0-1)$ and their reduction with respect to $\mathrm{SU}(2,2)\times\mathrm{U}(1)$ is given by[17]

$$D_S(s_0+1;0;s_0;-s_0-1)$$

$$\underset{\mathrm{SO}(4,2)\times\mathrm{U}(1)}{\Longrightarrow} D(s_0+1;0,s_0)\times[-s_0-1]$$

$$\oplus D(s_0+\tfrac{3}{2};0,s_0)\times[-s_0+\tfrac{1}{2}]. \tag{4.4'}$$

The reduction of these representations with respect to Osp(1,4) and SO(3,2) is the same as the reduction of $D_S(s_0+1;s_0,0;s_0+1)$, i.e., they contain the irrep $D_S(s_0+1,s_0)$ of Osp(1,4), which reproduces with respect to its bosonic part according to (4.5). They differ, however, by the $\mathrm{U}(1)_Y$ content and by the SO(4) content.

To explain this we consider the two Casimir operators of $SO(4)_{S_{i5},S_i}$,

$$\mathscr{C}^{(1)}(SO(4)) \equiv \tfrac{1}{2} S_{\alpha\beta} S^{\alpha\beta} \quad (\alpha\beta \text{ summed over } 1,2,3,5),$$
$$\mathscr{C}^{(2)}(SO(4)) \equiv S_{i5} S_i \quad (i \text{ summed over } 1,2,3). \tag{4.13}$$

From relation (4.6), it follows after some calculations that they can take the following values:

$$\mathscr{C}^{(1)}(SO(4)) = \Gamma_0^2 + s^2 - 1 = \mu^2 + s^2 - 1, \tag{4.14}$$

$$\mathscr{C}^{(2)}(SO(4)) = +s\Gamma_0 = +s\mu \tag{4.15a}$$

or

$$\mathscr{C}^{(2)}(SO(4)) = -s\Gamma_0 = -s\mu. \tag{4.15b}$$

The sign of the eigenvalue of $\mathscr{C}^{(2)}(SO(4))$ is an invariant of the representations. It distinguishes between the two representations characterized by the same value of s. Equation (4.15a) holds in $D(s+1;s,0)$ and therefore in $D_S(s_0+1;s_0,0;s_0+1)$, and (4.15b) holds in $D(s+1,0;s)$ and therefore in $D_S(s_0+1;0,s_0;-s_0-1)$.[20]

Equation (4.15) gives us the meaning of the quantum number s which will lead to its physical interpretation: s is, according to (4.13), the component of angular momentum S_i along $S_{i5}\Gamma_0^{-1}$. In the one irrep this component is positive, in the other it is negative.

There are now two possibilities for the extension of SU(2,2/1) by the parity operator. We will assume that $S_{\mu\nu}$ has the usual parity property of Lorentz generators $U_p S_{ij} U_p = S_{ij}$, $U_p S_{0i} U_p = -S_{0i}$ and Γ_μ is a Lorentz vector

$$U_p \Gamma_\mu U_p = \epsilon(\mu)\Gamma_\mu;$$
$$\epsilon(\mu) = \begin{cases} +1, & \text{for } \mu = 0, \\ -1, & \text{for } \mu = i = 1,2,3. \end{cases}$$

These two possibilities are the following.

Case (1): S_{i5} is a vector, S_{05} a scalar,

$$D = S_{65} \text{ a scalar,} \tag{4.16a}$$

and the Majorana spinors Q, S and the U(1) generator Y have the following transformation properties:

$$U_p Q U_p = \eta A Q, \quad U_p S U_p = \eta A S, \quad U_p Y U_p = -Y. \tag{4.16b}$$

Case (2): S_{i5} is a pseudovector, S_{05} a pseudoscalar,

$$S_{65} \text{ a pseudoscalar,} \tag{4.17a}$$

and Q, S, and Y have the following properties:

$$U_p Q U_p = \eta' A S, \quad U_p S U_p = \eta'^* A Q, \quad U_p Y U_p = Y. \tag{4.17b}$$

Here η and η' are phase factors and A is the standard matrix with the property $A\gamma_\mu A^{-1} = \gamma_\mu^\dagger$.

In the interpretation where $\mathfrak{P}_\mu = S_{\mu 5} + \Gamma_\mu$ are the momenta, only case (1) is possible. In the usual four-dimensional representation with $S_{\mu 6} = \Gamma_\mu = \frac{1}{2}\gamma_\mu$, $S_{\mu 5} = (i/4)\gamma_\mu\gamma_5$, $S_{56} = \gamma_5$ only case (2) is possible. As we will use neither of these representations, both cases are still possible.

We will denote the spaces for the representations $D_S(s_0+1;0,s_0;-s_0-1)$ and their basis vectors by symbols analogous to (4.9)–(4.12) with $+s_0$ now replaced by $-s_0$ and $+s$ by $-s$.

In case (1), U_p transforms from $\mathfrak{H}^{+s_0}$ to $\mathfrak{H}^{-s_0}$ and from $\mathfrak{H}^{+s}$ to $\mathfrak{H}^{-s}$; the vectors $|\mu, jj_3, +s_0\rangle$ and $|\mu, jj_3 - s_0\rangle$ are not parity eigenvectors. An irrep space of SU(2,2/1) extended by parity is given by the direct sum

$$\mathfrak{H}^{+s_0} \oplus \mathfrak{H}^{-s_0}$$

and parity eigenvectors are given by the linear combinations

$$|\mu, j, j_3; \pm\rangle = (1/\sqrt{2})(|\mu, j, j_3; +s\rangle \pm |\mu, j, j_3; -s\rangle). \tag{4.18}$$

One can arrange arbitrary phase factors (intrinsic parity and η) such that

$$U_p|\mu, j, j_3, \pm\rangle = \pm(-1)^{[j]}|\mu, j, j_3, \pm\rangle, \tag{4.19}$$

where $[j]$ = largest integer equal to or smaller than j.

In this case (parity doubling) the vector operator S_{i5}/Γ_0 specifies a direction inside the extended object (with noncommuting components) and s (or $-s$) is the component of the angular momentum S_j along this direction. One has a situation very similar to that of a quantum mechanical dumbbell with a flywheel on its axis (Ref. 12, p. 198) only that now the operator which specifies this axis, S_{i5}, has noncommuting components. Physical states can be either parity eigenstates (hadron resonances as vibrational and rotational excitations of an extended object) or eigenstates of $S_{i5}S_i$ (charged particle in the field of a dyon with electric charge and magnetic charge s/e).[10]

In case (2), U_p does not transform out of the space $\mathfrak{H}^{+s_0}$ or $\mathfrak{H}^{-s_0}$ and the $|\mu j j_3 s_0\rangle$ can be made parity eigenstates with parity

$$U_p|\mu j j_3; s_0\rangle = (-1)^{[\mu - 1]}|\mu j j_3; s_0\rangle. \tag{4.20}$$

In this case the spaces $\mathfrak{H}^\mu(j)$ can represent hadron resonances without having to resort to parity doubling, except that the interpretation of the now axial vector S_{i5} is not as straightforward as in case (1).

The physical meaning of the quantum number s [which can take the two values s_0 and $s_0 + \frac{1}{2}$ in an irrep of SU(2,2/1)] depends upon the realization of the SU(2,2/1) representations. If the generators of SU(2,2) represent the observables of a dyon system then s/e is the magnetic charge.[10] For this realization only case (1) is possible. If the generators of SU(2,2/1) represent the observables for the intrinsic motion of a relativistic oscillator,[21] then s represents the total spin of the constituents. Applying this relativistic oscillator to the hadrons would therefore require that $s = \frac{1}{2}$ (for baryons with the sum of the quark spin $\frac{1}{2}$) and $s = 1$ (for the mesons with the sum of the quark spins equal 1), which means that we have to choose the irrep with $s_0 = \frac{1}{2}$. This irrep can therefore be used to explain the equality of the slope for the meson and baryon trajectories[22] by combining them into an infinite supermultiplet of vibrational and rotational excitations.

In order to apply the SU(2,2/1) to hadrons whose center of mass (c.m.) motion is described by the Poincaré group $\mathscr{P}$ one has to couple the SU(2,2/1) describing the intrinsic motion with the $\mathscr{P}$ describing the motion of the hadron as a whole (c.m. motion). This coupling is done using constrained Hamiltonian quantum mechanics as we shall discuss now.

V. COUPLING SU(2,2/1) TO THE POINCARÉ GROUP

The representations discussed in the previous sections provide the frame for physical models, they do not yet fix the theory completely.

In nonrelativistic quantum mechanics one specifies the model by the choice of the Hamiltonian. For example, one obtains the harmonic oscillator[6] if one chooses in $\mathfrak{H}^{\mathrm{Osp}(1,2)}$ for the Hamiltonian

$$H^{\text{intrinsic}} = k\Gamma_0 + \text{const}$$
$$= (k/4)(\{Q,Q^\dagger\} + \{S,S^\dagger\}) + \text{const} \quad (5.1)$$

[the second equality in (5.1) follows from the defining relations (3.3) of Osp(2,2)]. In the nonrelativistic case the center of mass (c.m.) motion is usually ignored but the total Hamiltonian is actually the c.m. Hamiltonian $\mathbf{P}^2/2m$ plus the intrinsic Hamiltonian $H^{\text{intrinsic}}$

$$H = \mathbf{P}^2/2m + k\Gamma_0 \quad (5.2)$$

and the space of physical states is $\mathfrak{H} = \mathfrak{H}^{\text{c.m.}} \times \mathfrak{H}^{\mathrm{Osp}(1,2)}$, which is spanned by $|\mathbf{p}\rangle \otimes |q_0;q,\mu\rangle$, where $|\mathbf{p}\rangle$ are the generalized eigenvectors of $\mathbf{P}$ (plane wave of the c.m.) and $|q_0;q,\mu\rangle$ are the basis vectors of (2.3). For each $\mathbf{p}$ one has the same algebra $\mathrm{SU}(1,1/1)_{\Gamma_0,B_\pm,Q,S}$.

In the relativistic case the combination of c.m. and intrinsic motion is not trivial. The specific model is again defined by the choice of the Hamiltonian. If one wants to consider the relativistic supersymmetric oscillator one chooses the relativistic Hamiltonian as

$$H = v\Big(P_\mu P^\mu - \frac{1}{\alpha'}\frac{1}{16}\sum_{\beta=1}^{4}\big(\{\hat{Q}_\beta,\hat{Q}^\dagger_\beta\}$$
$$+ \{\hat{S}_\beta,\hat{S}^\dagger_\beta\}\big) + M_0^2\Big)$$
$$= v(P_\mu P^\mu - (1/\alpha')\hat{P}_\mu\Gamma^\mu - m_0^2) \quad (5.3)$$

[the second equality in (5.3) follows from the defining relations (3.3) of SU(2,2/1)].

In (5.3) P_μ is the c.m. momentum, $\hat{P}_\mu = P_\mu M^{-1}$ where $M^2 = P_\mu P^\mu$ and v is the Lagrange multiplier of constraint Hamiltonian mechanics which becomes $-(1/2M)$ when the evolution parameter is the proper time of the center of mass.

The $\hat{Q}_\alpha$ and $\hat{S}_\alpha$ in (5.3) are the components of the Majorana spinor of an SU(2,2/1) defined by (3.3). This SU(2,2/1) describes—like the Su(1,1/1) for the nonrelativistic case—the intrinsic motion. The c.m. motion is now described by the Poincaré group $\mathscr{P}_{P_\mu J_{\mu\nu}}$ $\mathrm{SO}(3,1)_{J_{\mu\nu}}$. Unlike the nonrelativistic case the combination of c.m. motion and intrinsic motion is not given by the direct product of spaces. Therefore one does not have any more one and the same algebra SU(2,2/1) for each momentum P_μ. However, one can construct a representation space such that for every eigenvalue $(P_\mu/m) = \hat{P}_\mu$ of the operator $\hat{P}_\mu = P_\mu M^{-1}$ one has an $\mathrm{SU}(2,2/1)_{\hat{S}_{\mu\nu}\hat{\Gamma}_\mu\hat{S}_{5_\mu}\hat{Q}_\alpha\hat{S}_\alpha}$, where the generators $\hat{\Gamma}_\mu$, $\hat{Q}_\alpha$, etc., depend upon $\hat{P}_\mu$.[23] In particular the operators that appear in (5.3) are defined in terms of the original generators of the $\mathrm{SU}(2,2/1)_{S_{\mu\nu}\Gamma_\mu S_{5_\mu}Q_\alpha S_\alpha}$ in the rest frame by

$$\hat{Q}_\alpha = \hat{Q}_\alpha(\hat{P}) = U^{-1}(L)Q_\alpha U(L) = D(\hat{P})_{\alpha\beta}Q_\beta \quad (5.4)$$

and similar expressions for the other generators, e.g.,

$$\hat{\Gamma}^0(\hat{P}) = U^{-1}(L)\Gamma^0 U(L) = L^0_\mu\Gamma^\mu = \hat{P}_\mu\Gamma^\mu. \quad (5.4')$$

Here $L^\mu_\nu = L^\mu_\nu(\hat{P})$ denotes the operator matrix for the inverse boost,

$$L(\hat{P})^\mu_\nu = \begin{pmatrix} \hat{P}^0 & \hat{P}_m \\ -\hat{P}^m & \delta^m_n - \hat{P}^m\hat{P}_n(1+\hat{P}_0)^{-1} \end{pmatrix},$$
$$\mu,\nu = 0,1,2,3, \quad m,n = 1,2,3, \quad (5.5)$$

which has the property

$$L(\hat{P})^\mu_\nu\hat{P}^\nu = \eta^{\mu 0}1.$$

The $D(\hat{P})$ is the 4×4 matrix that represents the Lorentz transformation $L(\hat{P})$ in the spinor representation

$$D(\hat{p}) = \sqrt{1/2(1+\hat{p}_0)}\,(1 + \hat{P}_\mu\gamma^\mu\gamma^0). \quad (5.6)$$

The basis vectors that span the representation space are now

$$|m;\hat{p}_i,\mu,j,j_3;s\rangle = \mathrm{U}(L^{-1}(\beta))(|m,\hat{p}_i = 0\rangle$$
$$\otimes|\mu,j,j_3,s\rangle), \quad (5.7)$$

where $|\mu,j,j_3;s\rangle$ are the basis vectors of (4.9) and $|m,\hat{p}_i = 0\rangle$ are the Wigner basis vectors at rest for the representation $(m, j=0)$ of the Poincaré group. One can show[23] that the $|m;\hat{p}_i,\mu,j,j_3;s\rangle$ are eigenvectors of $\hat{P}_\mu\Gamma^\mu$, $\hat{W} = -\hat{w}_\mu\hat{w}^\mu$ ($\hat{w}_\mu = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\hat{P}^\nu J^{\rho\sigma}$) with eigenvalue μ and $j(j+1)$, respectively, and that they transform under a Lorentz transformation $\mathrm{U}(\Lambda)$ like the Wigner basis vectors,

$$\mathrm{U}(\Lambda)|m;\hat{p}_i,\mu\, j\, j_3;s\rangle$$
$$= \sum_{j_3'}|m;(\Lambda\hat{p})_i\mu,j,j_3's\rangle D^{(s)}_{j_3'j_3}(\mathscr{R}), \quad (5.8)$$

where $\mathscr{R}$ is the Wigner rotation $\mathscr{R} = L(\Lambda\hat{p})\Lambda L^{-1}(\hat{p})$. The $\hat{Q}_\alpha$, $\hat{S}_\alpha$, $\hat{\Gamma}_i$, etc., act on the $|m;\hat{p}_i,\mu,j,j_3;s\rangle$ in the same way as the Q_α, S_α, Γ_i, etc., act on the $|\mu,j,j_3;s\rangle$.

A representation space for which (5.7) are the basis vectors can be constructed for every value of m^2 = eigenvalue $P_\mu P^\mu > 0$ leading to the space of off-mass-shell states. One can then put this relativistic system on the oscillator mass shell by applying the constraint[24]

$$H/v \approx 0.$$

For the relativistic harmonic oscillator defined by the Hamiltonian (5.3) this means

$$P_\mu P^\mu - (1/\alpha')\hat{P}_\mu\Gamma^\mu - m_0^2 \approx 0 \quad (5.9)$$

or

$$m^2 = m_0^2 + (1/\alpha')\mu. \quad (5.9')$$

The mass thus becomes a function of the internal quantum numbers μ, j, s; for the particular case of (5.3) its dependence upon j and s is trivial and it only depends upon the vibrational quantum number μ.

To each level (μ, j) in Fig. 2 there corresponds now an irreducible representation space $\mathfrak{H}^\mu(m,j)$ of the Poincaré group where $m^2 = m^2(\mu,j,s)$, e.g., as given by (5.9'). In place of (4.12) with (4.10) and (4.11) we have now (for $s_0 = \frac{1}{2}$)

$$\mathfrak{H}^{+1/2} = \sum_{\substack{\mu = 3/2,5/2,\ldots \\ j = 1/2,3/2,\ldots,\mu - 1}} \oplus \mathfrak{H}^{s = 1/2,\mu}(m;j)$$

$$\oplus \sum_{\substack{\mu = 2,3,4,\ldots \\ j = 1,2,3,\ldots,\mu - 1}} \oplus \mathfrak{H}^{s = 1,\mu}(m;j). \qquad (5.10)$$

The first term describes the baryon resonances that lie on a Regge trajectory (states with $\mu = j + 1$ in Fig. 2) and their daughters ($\mu > j + 1$). The second term in (5.10) describes the meson resonances.

The mass formula (5.9′) predicts then the equality of the slope for the meson and baryon Regge trajectories.[22] It does not yet give the correct mass formula, which, however, can be easily obtained if one adds to the Hamiltonian (5.3) a term whose eigenvalue is $(1/\alpha')s^2$. There are many candidates for such a term, e.g., the second-order Casimir operator of the SO(3,2) subgroup, or the square of $S_i S_{i5} \Gamma_0^{-1}$ (component of angular momentum along the "direction" S_{i5}).

The choice of the Hamiltonian (5.3) is as arbitrary as the choice of (5.1) or the choice of any Hamiltonian. The real question is whether there is a physical system which is—at least approximately—described by the chosen Hamiltonian. So we are free to modify (5.3) according to the needs of the experiments to get relativistic vib-rotors, etc. That the system described with the Hamiltonian (5.3) was called a harmonic oscillator can only be justified by correspondence.[21] A nonrelativistic limit gives, for $s = 0$, the isotropic oscillator in three dimensions.

[1] E. Witten, Nucl. Phys. B **188**, 513 (1981).

[2] S. Goshen and H. J. Lipkin, Ann. Phys. (NY) **6**, 301 (1959); A. O. Barut and A. Bohm, Phys. Rev. **139**, B1107 (1965); Y. Dothan, M. Gell-Mann, and Y. Ne'eman, Phys. Lett. **17**, 145 (1965). SGG will also be called dynamical group.

[3] V. de Alfaro, S. Fubini, and G. Furlan, Nuovo Cimento A **34**, 569 (1976); R. Jackiw, Ann. Phys. (NY) **129**, 183 (1980).

[4] S. Fubini and E. Rabinovici, Nucl. Phys. B **245**, 17 (1984); E. D'Hoker and L. Vinet, Phys. Lett. B **137**, 72 (1984).

[5] B. G. Adams, J. Cizek, and J. Paldus, "Lie algebraic methods and their applications to simple quantum systems," in *Dynamical Groups and Spectrum Generating Algebras* (World Scientific, Singapore, to be published).

[6] A. B. Balantekin, Ann. Phys. (NY) **164**, 277 (1985); J. Beckers, D. Dehin, and V. Hussin, J. Phys. A **20**, 1137 (1987).

[7] E. D'Hoker and L. Vinet, Commun. Math. Phys. **97**, 391 (1985).

[8] A. H. Kamupingene, T. D. Palev, and S. P. Tsaneva, International Center for Theoretical Physics, Trieste, preprint IC/85/169.

[9] V. A. Kostelecky, M. M. Nieto, and D. R. Truax, Phys. Rev. D **32**, 2627 (1985).

[10] E. D'Hoker and L. Vinet, Nucl. Phys. B **260**, 79 (1985); A. O. Barut and J. L. Bornzin, J. Math. Phys. **12**, 841 (1971).

[11] M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146 (1977).

[12] A. Bohm, *Quantum Mechanics—Foundations and Applications* (Springer, New York, 1986), 2nd ed. Page 8 gives the definition of * algebra and the remainder of the book shows that quantum mechanics needs star and not grade-star algebras.

[13] C. Fronsdal and T. Mirai, "Unitary representations of supergroups," in *Essays on Supersymmetry* (Reidel, Boston, 1986).

[14] V. Bargmann, Ann. Math. **48**, 568 (1947). $D_+(k)$ in Bargmann's notation is our $D_+(q)$.

[15] For the finite dimensional case (grade-star but not star representations) they have been given in M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977); M. Baake and P. Reinicke, *ibid.* **27**, 859 (1986). For infinite dimensional representations see also J. W. B. Hughes, J. Math. Phys. **22**, 245 (1981).

[16] For SU(2,2/1) we use the conventions of M. F. Sohnius, Phys. Rep. **128**, 182 (1985), except that M_{ab} there is called S_{ab} here. We also use different symbols $\mathfrak{P}_\mu$, $\mathfrak{K}_\mu$ in place of the usual P_μ and K_μ in order to emphasize that these operators in our physical interpretation are not the usual generators of translations and conformal transformations but, like the other generators of our SU(2,2/1), observables connected with the intrinsic dynamics of an extended relativistic object. The symbols P_μ we shall reserve for the translation generators of the center of mass motion of our "relativistically vibrating" extended object.

[17] V. K. Dobrev and V. B. Petkova, Phys. Lett. B **162**, 127 (1985); M. Flato and C. Fronsdal, Lett. Math. Phys. **8**, 159 (1984); C. Fronsdal, Phys. Rev. D **26**, 1988 (1982); W. Heidenreich, Phys. Lett. B **110**, 461 (1982); D. Z. Freedman and H. Nicolai, Nucl. Phys. B **237**, 342 (1984).

[18] G. Mack, Commun. Math. Phys. **55**, 1 (1977); G. Mack and I. T. Todorov, J. Math. Phys. **10**, 2078 (1969). Here $j^{(1)}$ and $j^{(2)}$ are related to the highest value of angular momentum, $j^{\max} \equiv k_1$, and the lowest value of angular momentum, $j^{\min} = k_0$, by $k_0 = |j^{(1)} - j^{(2)}| \pm (k_1 + 1) = (j^{(1)} + j^{(2)} + 1)\mathrm{sign}(j^{(1)} - j^{(2)})$.

[19] A. O. Barut and A. Bohm, J. Math. Phys. **11**, 2938 (1970).

[20] Similar relations follow from (4.6) for the noncompact subgroups $SO(3,1)_{S_{i0}S_i}$ and $SO(1,1)_{S_{56}}$:

$$\mathscr{C}^{(2)}(SO(3,1)) \equiv S_{i0} S_i = +sS_{56} \quad \text{in } D(s+1;s,0)$$

$$\mathscr{C}^{(2)}(SO(3,1)) \equiv -sS_{56} \quad \text{in } D(s+1;0,s).$$

[21] A. Bohm, M. Loewe, P. Magnollay, L. C. Biedenharn, H. van Dam, M. Tarlini, and R. R. Aldinger, Phys. Rev. D **32**, 2828 (1985).

[22] A. Bohm, Phys. Rev. Lett. **57**, 1203 (1986).

[23] A. Bohm, M. Loewe, L. C. Biedenharn, and H. van Dam, Phys. Rev. D **28**, 3032 (1983). See also D. J. Almond, J. Phys. G **9**, 1309 (1983), where the same construction has been given for the relativistic string.

[24] A. J. Hanson and T. Regge, Ann. Phys. (NY) **87**, 498 (1974); N. Mukunda, H. van Dam, and L. C. Biedenharn, *Relativistic Models of Extended Hadrons Obeying a Mass-Spin Trajectory Constraint* (Springer, New York, 1982).

Threshold singularities of two-cluster–two-cluster scattering amplitudes for dilation analytic potentials

Jan Dereziński
Division of Mathematical Methods in Physics, Warsaw University, Hoża 74, 00-682 Warsaw, Poland

(Received 18 December 1986; accepted for publication 16 December 1987)

Two-cluster–two-cluster scattering amplitudes are studied for N-body quantum systems with potentials that are both dilation analytic and exponentially decaying. It is proved that under quite broad assumptions these amplitudes can be meromorphically continued in the energy, with square root or logarithmic branch points at the two-cluster thresholds.

I. INTRODUCTION

Two-cluster–two-cluster scattering amplitudes are usually easier to study than other kinds of many-body scattering amplitudes. In Ref. 1, we studied their analytic properties below the lowest three-cluster threshold. We proved there that if the potentials decay exponentially then the two-cluster–two-cluster scattering amplitudes can be in this energy range continued meromorphically with square root or logarithmic branch points at the two-cluster thresholds. In this paper we study analogous properties of those amplitudes in the whole energy range.

To our knowledge, there are two distinct approaches that have been used in the study of analytic properties of scattering amplitudes. The approach that can be found in Refs. 1 and 2 is based on some special resolvent equation. The other approach[2–8] relies on the dilation analyticity technique. By this technique one can prove the following. Suppose that the potentials both decay exponentially and are dilation analytic, and the channel eigenvectors have a sufficient decay. Then the two-cluster–two-cluster scattering amplitudes can be continued meromorphically in a neighborhood of the real axis outside of the threshold (see especially Refs. 5 and 8).

In our paper we join those two approaches. We make the same assumptions as in the second approach. We prove that the two-cluster–two-cluster scattering amplitudes can be continued meromorphically around all the two-cluster thresholds that are not simultaneously more-than-two-cluster thresholds. At those thresholds we obtain square root branch points if the dimension is odd and logarithmic branch points if the dimension is even, which is the same behavior as in the two-body case.[1,9]

II. NOTATION

In this paper we use the notation from Ref. 1. For the reader's convenience, we will give a short summary of this notation.

We study a many-body Schrödinger operator acting on $L^2(\mathbf{R}^{dN})$ defined by

$$H = -\sum_{i=1}^{N} \frac{\Delta_i}{2m_i} + \sum_{\substack{i,j=1\\ i<j}}^{N} V_{ij}(x_i - x_j).$$

In a standard way we remove the center of mass motion, introduce the concepts of a cluster decomposition D, a cluster Hamiltonian H_D, etc.[1,9–11]

The variables x^D stand for intracluster degrees of freedom and the x_D denote intercluster degrees of freedom. If we represent the original Hilbert space as the tensor product $L^2(x^D) \otimes L^2(x_D)$ we can write our cluster Hamiltonian as

$$H_D = H^D \otimes 1 + 1 \otimes T_D,$$

where T_D is the kinetic energy of the c.m. motion of the clusters.

If D is an i-cluster decomposition with $1 < i < N$, then eigenvalues of H^D are called i-cluster thresholds. The point zero is the only N-cluster threshold. The elements of $L^2(X^D)$ that are eigenvectors of H^D we denote by ϕ_α and call channels. We denote the threshold corresponding to the channel ϕ_α by ω_α and the corresponding cluster decomposition by $D(\alpha)$. If $D(\alpha)$ is a two-cluster decomposition then the corresponding reduced mass of the intercluster motion we denote by μ_α and $\nu_\alpha(z)$ will stand for $(2\mu_\alpha(z-\omega_\alpha))^{1/2}$. The generalized eigenvector of $H_{D(\alpha)}$ corresponding to the channel ϕ_α with the intercluster momentun k we denote by $\phi_\alpha(k)$, explicitly:

$$\phi_\alpha(k)(x) = \phi_\alpha(x^{D(\alpha)})\exp(ikx_{D(\alpha)}).$$

We define also

$$T_\alpha = \omega_\alpha + T_{D(\alpha)} = \omega_\alpha - \Delta_{D(\alpha)}/2\mu_\alpha.$$

The scattering amplitude for the α-β scattering at energy λ is given by the formula

$$\begin{aligned} t_{\alpha\beta}(k_1,k_2) &= (\phi_\alpha(k_1),(V - V_{D(\alpha)})\phi_\beta(k_2)) \\ &\quad + \lim_{\epsilon\to 0^+}(\phi_\alpha(k_1),(V - V_{D(\alpha)}) \\ &\quad \times R(\lambda + i\epsilon)(V - V_{D(\beta)})\phi_\beta(k_2)), \end{aligned}$$

where $T_\alpha\phi_\alpha(k_1) = \lambda\phi_\alpha(k_1)$, $T_\beta\phi_\beta(k_2) = \lambda\phi_\beta(k_2)$, and $(\cdot)$ denotes the scalar product.[5,13]

We denote by P_D the orthogonal projection onto the point spectrum of H^D. Also $|x|$ means some fixed Euclidean norm of the vector x. The symbols ρ^b, $\rho^{D,b}$, and $\rho_D{}^b$ will denote the operators of multiplication by $\exp(-b(|x|^2+1)^{1/2})$, $\exp(-b(|x^D|^2+1)^{1/2})$, and $\exp(-b(|x_D|^2+1)^{1/2})$, respectively.

In contrast to Ref. 1, in this paper we will use the dilation analyticity technique. So we have to introduce a family of unitary operators on $L^2(\mathbf{R}^d)$ defined by $(\Gamma(\theta)f)(x) = \exp(d\theta/2)f(\exp(\theta)x)$.

 0022-2488/88/051171-03$02.50

III. MAIN RESULTS

The assumption that we impose on the potentials is a kind of unification of the dilation analyticity condition[13,14] and of the condition of an exponential decay. It appeared before in a similar situation.[4,5,8]

Assumption 3.1: The potentials V_{ij} can be factored into $W_{ij}^{(1)} W_{ij}^{(2)}$ such that for some $\epsilon > 0$ and $\gamma > 0$, $\Gamma(\theta) W_{ij}^{(k)} \Gamma(-\theta)(\rho^{ij,c})^{-1}(1-\Delta^{ij})^{-1/2}$ extends to analytic families in $|\mathrm{Im}\,\theta| < \gamma$ with values in compact operators.

For simplicity we will conduct parts of our reasoning under a stronger assumption.

Assumption 3.2: For some $c > 0$ and $\gamma > 0$, $\Gamma(\theta) V_{ij} \Gamma(-\theta)(\rho^{ij,2c})^{-1}$ extends to an analytic family in $|\mathrm{Im}\,\theta| < \gamma$ with values in L^∞.

It is easy to extend our proof to cover singular potentials satisfying Assumption 3.1 by following the ideas of Sec. V of Ref. 1.

Now we state the main theorem of this section.

Theorem 3.3: Suppose Assumption 3.1 holds. Fix two unit vectors $\hat{e}_1$ and $\hat{e}_2$ and two channels ϕ_α and ϕ_β that correspond to two-cluster decompositions $D(\alpha)$ and $D(\beta)$, respectively. Assume that the threshold energies ω_α and ω_β corresponding to these channels do not coincide with thresholds of $H_{D(\alpha)}$ and $H_{D(\beta)}$, respectively. Then the following is true.

(a) The scattering amplitude $t_{\alpha\beta}[\nu_\alpha(z)\hat{e}_1, \nu_\beta(z)\hat{e}_2]$ exists and can be extended to a meromorphic function of z in a neighborhood of the real axis outside the thresholds of H.

(b) If ω is a two-cluster threshold that does not coincide with a more-than-two-cluster threshold and d is odd then $t_{\alpha\beta}[\nu_\alpha(z)\hat{e}_1, \nu_\beta(z)\hat{e}_2]$ can be extended to a meromorphic function of z in a neighborhood of ω on the Riemann surface of $(z-\omega)^{1/2}$. If d is even then the same is true except that $\log(z-\omega)$ replaces $(z-\omega)^{1/2}$.

Remark 3.4: (a) The result (a) has been known before (see Refs. 5 and 8).

(b) If we are interested just in the existence of the two-cluster–two-cluster amplitude we do not need to assume either the dilation analyticity or the exponential decay of the potentials. Instead we can apply the results of Ref. 15 and get the existence for a much wider class of potentials.

Now we present the main facts used in the proof of Theorem 3.3. Unless stated otherwise we will suppose Assumption 3.1 to be true.

Theorem 3.5: (Balslev and Combes[13,14]) The expression $\Gamma(\theta) H_D \Gamma(-\theta)$ defined for real θ can be extended to an analytic family for $|\mathrm{Im}\,\theta| < \gamma$. Let us denote this family by $H_D(\theta)$. The continuous spectrum of $H_D(\theta)$ is equal to

$$\bigcup_{\omega_i} \omega_i + \exp(2\theta)\mathbb{R}_+,$$

where ω_i runs over the set of all the threshold of H_D.

By $R(\theta,z)$, $\phi_\alpha(\theta)$, etc., we will denote the unique analytic continuation of $\Gamma(\theta)R(z)\Gamma(-\theta)$, $\Gamma(\theta)\phi_\alpha$, etc.

Lemma 3.6: Suppose that ω_α is not a threshold of $H^{D(\alpha)}$. Then for some $a > 0$ and $\gamma' > 0$ the vector $(\rho^{D(\alpha),a})^{-1} \times \phi_\alpha(\theta)$ belongs to L^2 uniformly for $|\mathrm{Im}\,\theta| < \gamma'$.

Proof: See Theorem XII.41 of Refs. 14 and 16. Q.E.D.

The next Lemma follows easily by the standard dilation analyticity techniques.[13,14]

Lemma 3.7: Suppose $0 < \mathrm{Im}\,\theta < \gamma$. Let ω be a two-cluster threshold that does not coincide with a more-than-two-cluster threshold. Then (a) ω is isolated in the set of the thresholds of H; and (b) the following functions are analytic in z in a neighborhood of ω: $R_D(\theta,z)(1-P_D(\theta))$ for a two-cluster decomposition D, and $R_D(\theta,z)$ for a more-than-two-cluster decomposition D.

Lemma 3.8: Under the same assumption as above, if $b > 0$, d is odd, and D is a two-cluster decomposition, then the operator-valued function $\rho^{D,b} R_D(\theta,z) P_D(\theta) \rho^{D,b}$ defined for all z in a neighborhood of ω, except maybe for the cut $\omega + \exp(2\theta)\mathbb{R}_+$, can be extended analytically onto a neighborhood of ω on the Riemann surface of the function $(z-\omega)^{1/2}$. If d is even the same is true except that $\log(z-\omega)$ replaces $(z-\omega)^{1/2}$.

Proof: Note that

$$R_D(\theta,z)P_D(\theta) = \sum_{D(\alpha)=D} P_\alpha(\theta) \otimes (z - \omega_\alpha - \exp(2\theta)T_D)^{-1},$$

where P_α is the projection corresponding to a channel α and ω_α is its respective threshold. In the above sum all the terms with $\omega_\alpha \neq \omega$ are analytic around ω. The terms with $\omega_\alpha = \omega$ are treated as in the Appendix to §XI.6 of Ref. 9. Q.E.D.

Proof of Theorem 3.3: We are only going to outline the argument since its basic ingredients are already given in detail in Refs. 1 and 5.

We use the formula for $t_{\alpha\beta}$ and Eq. (3.8) of Ref. 1 and obtain

$$\begin{aligned} t_{\alpha\beta}(k_1,k_2) &= (\Phi_\alpha(k_1),(V - V_{D(\alpha)})\Phi_\beta(k_2)) \\ &\quad + \lim_{\epsilon\to 0^+} (\rho^{-1}(V - V_{D(\alpha)})\Phi_\alpha(k_1), \\ &\quad \rho[A(\lambda+i\epsilon)](1 - M(\lambda+i\epsilon))^{-1}B(\lambda+i\epsilon) \\ &\quad + C(\lambda+i\epsilon)]\rho\rho^{-1}(V - V_{D(\beta)})\Phi_\beta(k_2)). \end{aligned}$$

The first term on the right-hand side is easy to handle; we will concentrate on the second one. We fix $\epsilon > 0$. The function

$$\begin{aligned} &\theta \to (\rho^{-1}(\theta)(V(\theta) - V_{D(\alpha)}(\theta))\Phi_\alpha(\theta,k_1), \\ &\rho(\theta)[A(\theta,\lambda+i\epsilon)(1 - M(\theta,\lambda+i\epsilon))^{-1}B(\theta,\lambda+i\epsilon) \\ &\quad + C(\theta,\lambda+i\epsilon)]\rho(\theta)\rho^{-1}(\theta)(V(\theta) \\ &\quad - V_{D(\beta)}(\theta))\Phi_\beta(\theta,k_2)) \end{aligned}$$

does not depend on θ for θ real and is analytic for $|\mathrm{Im}\,\theta| < \gamma$; consequently it does not depend on θ at all. Now fix θ with $|\mathrm{Im}\,\theta| < \gamma$. Suppose for simplicity that Assumption 3.2 holds. By Lemmas 3.6–3.8 and methods from Secs. III and IV of Ref. 1,

$$\rho^{-1}(\theta)(V(\theta) - V_{D(\alpha)}(\theta))\Phi_\alpha(\theta,k_1),$$

$$\rho(\theta)A(\theta,\lambda+i\epsilon), \quad M(\theta,\lambda+i\epsilon),$$

$$B(\theta,\lambda+i\epsilon)\rho(\theta), \quad \rho(\theta)C(\theta,\lambda+i\epsilon)\rho(\theta),$$

and

$$\rho^{-1}(\theta)(V(\theta) - V_{D(\beta)}(\theta))\Phi_\beta(\theta,k_2)$$

can be extended to analytic functions on a neighborhood of ω on the Riemann surface of either $(z-\omega)^{-1}$ or $\log(z-\omega)$. Now we apply the analytic Fredholm theorem to the term $(1-M(\theta,\lambda+i\epsilon))^{-1}$, which completes our proof. Q.E.D.

[1] J. Dereziński, "Existence and analyticity of many-body scattering amplitudes at low energies," J. Math. Phys. **28**, 1080 (1987).
[2] E. Balslev, "Resonances in three-body scattering theory," Adv. Appl. Math. **5**, 260 (1984).
[3] E. Balslev, "Analytic scattering theory of quantum mechanical three-body systems," Ann. Inst. H. Poincaré A **32**, 125 (1980).
[4] G. Hagedorn, "A link between scattering resonances and dilation analytic resonances in few body quantum mechanics," Commun. Math. Phys. **65**, 181 (1979).
[5] I. M. Sigal (private communication).
[6] I. M. Sigal, "Scattering theory in many body quantum systems. Analyticity of the scattering matrix," in *Quantum Mechanics in Mathematics, Chemistry and Physics*, edited by K. E. Gustafson and W. P. Reinhardt (Plenum, New York, 1981), p. 307.
[7] I. M. Sigal, "Mathematical theory of single channel systems. Analyticity of scattering matrix," Trans. Am. Math. Soc. **270 N2**, 517 (1982).
[8] I. M. Sigal, "Analytic properties of the scattering matrix of many particle systems," Int. Eq. Oper. Theory **9**, 134 (1986).
[9] M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. III (Academic, New York, 1979).
[10] G. Hagedorn, "Asymptotic completeness for classes of two, three, and four particle Schrodinger operators," Trans. Am. Math. Soc. **258**, 1 (1980).
[11] I. M. Sigal, *Scattering Theory for Many-Body Quantum Mechanical Systems* (Springer, New York, 1983).
[12] R. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1982).
[13] E. Balslev and J. M. Combes, "Spectral properties of many-body Schrodinger operators with dilation analytic interactions, Commun. Math. Phys. **22**, 280 (1971).
[14] M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. IV (Academic, New York, 1978).
[15] P. Perry, I. M. Sigal, and B. Simon, "Spectral analysis of *N*-body Schrodinger operators," Commun. Math. Phys. **55**, 259 (1981).
[16] R. Froese and I. Herbst, "Exponential bounds and absence of positive eigenvalues for *N*-body Schrödinger operators," Commun. Math. Phys. **87**, 429 (1982).

Scattering-into-cones theorem for a certain impurity scattering

Sung Hyun Yoon
Department of Physics Education, Korea National University of Education, Chungwon Kun, Chungbuk 363-890, Korea

(Received 1 September 1987; accepted for publication 20 January 1988)

The scattering-into-cones theorem for impurity scattering in a periodically stratified potential in three dimensions is given.

I. INTRODUCTION

A simple and intuitive way of looking at the scattering of particles is to relate their motion in configuration space (at $t = +\infty$) to their momenta. Dollard's scattering-into-cones theorem[1] asserts the following intuitive fact: for a normalized free particle state $\Psi(x,t)$ (with the Hamiltonian $-\Delta$)

$$\lim_{t\to\infty}\int_C |\Psi(x,t)|^2\,d^3x = \int_C |\hat{\Psi}(k)|^2\,d^3k,$$

where C is an infinite cone with apex at the origin and $\hat{\Psi}(k)$ is the Fourier transform of Ψ. That is, the probability that in the infinite future the particle will be found in C is equal to the probability that its momentum lies in the same cone. The theorem remains true for interacting cases (with a natural modification of the right-hand side) and it was generalized to cases with more general free Hamiltonians.[2] Also, the theorem forms a justification for the usual relation between the scattering amplitude and the differential cross section.[3,4]

We generalize the scattering-into-cones theorem to the impurity scattering in a periodically stratified potential in three dimensions, i.e., to the scattering for the pair $(H_0 = -\Delta + V_p, H = H_0 + V)$ where V_p is a periodic potential (period 1) depending only on the first coordinate x_1 and V is a short range potential.

II. SCATTERING-INTO-CONES

The impurity scattering under the assumption that $V_p \in L^\infty(\mathbb{R})$ and $V \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}^3)$ was studied in Ref. 5. First, we recall that there exists a generalized Fourier transform which is a spectral representation for H_0 (see Properties II. 3 and 4 of Ref. 5).

Lemma 1: The unitary operator $U: L^2(\mathbb{R}^3_x) \to L^2(\mathbb{R}^3_k)$ given by

$$(Uf)(k) = \text{l.i.m.}\,\frac{1}{(2\pi)^{3/2}}\int \bar{\beta}(x_1,k_1) \times \exp(-ik_\perp\cdot x_\perp) f(x)\,d^3x, \quad f\in L^2(\mathbb{R}^3_x)$$

has the property $UH_0U^{-1} = M_{\lambda(k)}$ and

$$(U^{-1}g)(x) = \text{l.i.m.}\,\frac{1}{(2\pi)^{3/2}}\int \beta(x_1,k_1) \times \exp(ik_\perp\cdot x_\perp) g(k)\,d^3k, \quad g\in L_2(\mathbb{R}^3_k).$$

Here $\beta(x_1,k_1)$ is a suitably normalized Bloch solution of Hill's equation, i.e.,

$$-\beta''(x_1,k_1) + V_p(x_1)\beta(x_1,k_1) = \lambda_1(k_1)\beta(x_1,k_1),$$

$$\int_0^1 |\beta(x_1,k_1)|^2\,dx_1 = 1,$$

and

$$\beta(x_1,k_1) = \exp(ik_1x_1)\chi(x_1,k_1)$$

for some periodic (in x_1) function χ. Here $M_{\lambda(k)}$ is the maximal multiplication operator (in the quasimomentum space) by the energy function $\lambda(k) = \lambda_1(k_1) + k_\perp^2$.

Notation: We write $(Uf)(k) = \hat{f}(k)$.

As given in Ref. 5, as $|x|\to\infty$, $G_0(x,y:E+i0)$ (the Green's function for the Lippmann–Schwinger equation) have dominant contributions only from those k's on the constant energy surface Σ_E whose group velocities are parallel to the direction of observation, and we can expect an analog of the ordinary theorem of scattering into cones.

First we establish such a property for freely developing states (under H_0).

Property 1: Let C be a circular cone with apex at the origin, i.e., $C = \{x\in\mathbb{R}^3: x\cdot n \geqslant \alpha|x|\}$ where n is a unit vector and $0 < \alpha \leqslant 1$. Then, for any $\Phi \in L^2(\mathbb{R})^3$,

$$\lim_{t\to\infty}\int_C |\Phi(x,t)|^2\,d^3x = \int \chi_C(\nabla\lambda(k))|\hat{\Phi}(k)|^2\,d^3k, \tag{2.1}$$

where $\Phi(x,t) = (e^{-itH_0}\Phi)(x)$ and χ_C is the characteristic function of C.

Proof: It is easy to see that Eq. (2.1) is equivalent to the existence of the limit of a family of bounded operators in the weak operator topology. Thus it is sufficient to prove Eq. (2.1) on a dense subset of $L^2(\mathbb{R}^3)$, which we choose as

$$U^{-1}C_0^\infty(G) = \{\Phi\in L^2(\mathbb{R}^3);\ \hat{\Phi}\in C_0^\infty(G)\},$$

where

$$G = \mathbb{R}^3/\{k:\ k_1 = 0, \pm\pi, \pm 2\pi, \ldots \text{ or } \lambda_1''(k_1) = 0\}.$$

(See part II of Ref. 5 for special natures of the points $k_1 = n\pi$.) Since U is a spectral representation of H_0,

$$\Phi(x,t) = (e^{-iH_0t}\Phi)(x) = (2\pi)^{-3/2}\int e^{-i\lambda(k)t}\hat{\Phi}(k)\Psi(x,k)\,d^3k.$$

Note $\Psi(x,k) = e^{ik\cdot x}\chi(x_1,k_1)$, where $\chi(x_1+1,k_1) = \chi(x_1,k_1)$ and χ is piecewise smooth in k_1. For simplicity write $\Phi(x,t) = \Phi_t(x)$. Setting $\vartheta = x/t$ and defining $\Psi_t(\vartheta) = \Phi(\vartheta t,t)$, we have

 0022-2488/88/051174-03$02.50

$$\Psi_t(\vartheta) = (2\pi)^{-3/2} \int \hat{\Phi}(k) \times \exp(i(k\cdot\vartheta - \lambda(k))t)\chi(\vartheta_1 t, k_1) d^3k.$$

Since $\hat{\Phi}\in C_0^\infty(G)$ and the functions involved are smooth in a sufficiently small neighborhood of the support of $\hat{\Phi}$, we can apply the standard stationary phase method (see Bleinstein–Handelsman[6] for instance). Then,

$$\Psi_t(\vartheta) = \frac{t^{-3/2}}{2}\Big[\sum_j \exp(i(k_j\cdot\vartheta - \lambda(k_j))t) \times \hat{\Phi}(k_j)\chi(\vartheta_1 t, k_{j1}) \times \exp(\tfrac{1}{2}\pi i \operatorname{sgn}(\lambda_1''(k_j)))|\lambda_1''(k_{j1})|^{-1/2}\Big] + O(t^{-5/2}), \tag{2.2}$$

where $\nabla_k \lambda(k)|_{k=k_j} = \vartheta$ and the estimate of the remainder is uniform with respect to ϑ (since $\hat{\Phi}$ is of compact support). For each ϑ, there are only finitely many k_j's. Note that the stationary points can differ only in the first coordinates. Corresponding to the given cone C, let $C' = \{k\in\mathbb{R}^3: \nabla\lambda(k)\in C\}$ and define

$$\Phi_t^C(\vartheta t) = (2\pi)^{-3/2}\int_{C'} \hat{\Phi}(k)\exp(i(k\cdot\vartheta - \lambda(k))t) \times\chi(\vartheta_1 t, k_1) d^3k \equiv \Psi_t^C(\vartheta). \tag{2.3}$$

Let $\bar\vartheta$ be a fixed positive number and consider

$$F_C(\bar\vartheta,t) \equiv \int_{C,|x|<\bar\vartheta t} d^3x |\Phi_t^C(x)|^2 = \int_{C,|\vartheta|<\bar\vartheta} d^3\vartheta\, t^3 |\Psi_t^C(\vartheta)|^2.$$

From Eq. (2.2), as $t\to\infty$

$$F_C(\bar\vartheta,t) = \int_{C,|\vartheta|<\bar\vartheta} d^3\vartheta \Big| \sum_j \exp(i(k_j\cdot\vartheta - \lambda(k_j))t) \times \hat{\Phi}(k_j)\chi(\vartheta_1 t, k_{j1})\mathscr{G}(k_j) + O(t^{-1})\Big|^2, \tag{2.4}$$

where

$$\mathscr{G}(k_j) = \tfrac{1}{2}|\lambda_1''(k_{j1})|^{-1/2} \exp\left[\tfrac{1}{4}\pi i \operatorname{sgn}(\lambda_1''(k_{j1}))\right].$$

Observe that each $k_j(\vartheta)$ is real analytic in ϑ if ϑ is in a region where $k_{j1} \neq n\pi$, $\lambda_1''(k_{j1}) \neq 0$. Since $\chi(x_1,k_1)$ is periodic and C^1 in x_1, we can expand $\chi(x_1,k_1)$ in the uniformly convergent Fourier series

$$\chi(x_1,k_1) = \sum_n a_n(k_1)\exp(2\pi n x_1 i).$$

In view of the uniform continuity of $a_n(k_1)$ in the support of $\hat{\Phi}$, when we put this series into (2.4), we can do the integration term by term. Now we want the nonvanishing contribution to $F_C(\bar\vartheta,t)$ as $t\to\infty$. In the integral

$$F_C(\bar\vartheta,t) = \sum_{\substack{j,l\\ m,n}} \int_{C,|\vartheta|<\bar\vartheta} d^3\vartheta \{(\exp[it(\xi(k_j) - \xi(k_l) + 2\pi(n-m)\vartheta_1)]) \times(a_n(k_{j1})\bar{a}_m(k_{l1})\hat{\Phi}(k_j)\bar{\hat{\Phi}}(k_l)) \times\mathscr{G}(k_j)\bar{\mathscr{G}}(k_l))\} + O(t^{-1}),$$

where $\xi(k_j) = k_j\cdot\vartheta - \lambda(k_j)$, the nonvanishing contributions are only from the cases where the exponents are identically zero, because for the other cases we can use either the Riemann–Lebesgue lemma or the method of stationary phase depending on whether the ϑ-gradient of the exponent can or cannot vanish. Thus we have

$$\lim_{t\to\infty} F_C(\bar\vartheta,t) = \sum_{j,n}\int_{C,|\vartheta|<\bar\vartheta} |\mathscr{G}(k_j)|^2 \times|\hat{\Phi}(k_j)|^2 |a_n(k_{j1})|^2 d^3\vartheta = \sum_j \int_{C,|\vartheta|<\bar\vartheta} |\mathscr{G}(k_j)|^2 |\hat{\Phi}(k_j)|^2 d^3\vartheta,$$

since

$$\sum_n |a_n(k_{j1})|^2 = \int_0^1 |\chi_1(x_1,k_{j1})|^2 dx_1 = 1$$

(see Lemma 1). Now

$$|\mathscr{G}(k_j)|^2 = \frac{1}{4}|\lambda_1''(k_{j1})|^{-1} = J\left(\frac{\partial k_j}{\partial\vartheta}\right)$$

and $k_j(\vartheta)$'s have disjoint images locally. As a result,

$$\lim_{t\to\infty} F_C(\bar\vartheta,t) = \int_{C'(\bar\vartheta)} |\hat{\Phi}(k)|^2 d^3k, \tag{2.5}$$

where $C'(\bar\vartheta) = \{k\in\mathbb{R}^3: \nabla\lambda\in C \text{ and } |\nabla\lambda| < \bar\vartheta\}$. Writing

$$I_C(\bar\vartheta) = \int_{C'(\bar\vartheta)} |\hat{\Phi}(k)|^2 d^3k,$$

we have

$$I_C \equiv I_C(\infty) = \int_{C'(\infty)} |\hat{\Phi}(k)|^2 d^3k = \int_{C'} |\hat{\Phi}(k)|^2 d^3k > I_C(\bar\vartheta).$$

By the Parseval identity for U, Eq. (2.3) implies

$$I_C = \int |\Phi_t^C(x)|^2 d^3x.$$

Based on these, we claim

$$\lim_{t\to\infty}\int_C |\Phi_t^C(x)|^2 d^3x = I_C.$$

Clearly

$$0 < I_C - \int_C |\Phi_t^C(x)|^2 d^3x = (I_C - I_C(\bar\vartheta)) + (I_C(\bar\vartheta) - F_C(\bar\vartheta,t)) - \int_{C,|x|>\bar\vartheta t} |\Phi_t^C(x)|^2 d^3x.$$

Given $\epsilon>0$, choose a sufficiently large $\bar\vartheta$ such that

$$0 < I_C - I_C(\bar\vartheta) < \epsilon/3$$

and choose T_1 so that for $t > T_1$,

$$|I_C(\bar\vartheta) - F_C(\bar\vartheta,t)| < \epsilon/3 \quad [\text{see Eq. (2.5)}].$$

Also choose T_2 such that for $t > T_2$,

$$\lim_{t\to\infty}\int_{C,|\vartheta|>\bar\vartheta_t}|\Phi_t^C(x)|^2\,d^3x<\frac{\epsilon}{3}.$$

Thus for $t>\max(T_1,T_2)$,

$$\left|I_c-\int_C|\Phi_t^C(x)|^2\,d^3x\right|<\epsilon,$$

and we have proved the claim.

From this we also obtain the following properties.

(i) If C and $\overline{C}$ are disjoint cases, then

$$\lim_{t\to\infty}\int_{\overline{C}}|\Phi_t^C(x)|^2\,d^3x=0.$$

(ii) If $C\subset\widetilde{C}$, then

$$\lim_{t\to\infty}\int_C|\Phi_t^{\widetilde{C}}(x)|^2\,d^3x=\lim_{t\to\infty}\int_C|\Phi_t^C(x)|^2\,d^3x=I_C.$$

Property (i) is clear from the claim above (since the integrand is positive). For (ii), write $\widetilde{C}=C+\overline{C}$. Then, $\Phi_t^{\widetilde{C}}(x)=\Phi_t^C(x)+\Phi_t^{\overline{C}}(x)$ and

$$\int_C|\Phi_t^{\widetilde{C}}|^2\,d^3x=\int_C|\Phi_t^C|^2\,d^3x+\int_C|\Phi_t^{\overline{C}}|^2\,d^3x$$
$$+2\,\mathrm{Re}\int_C\Phi_t^C\,\overline{\Phi_t^{\overline{C}}}\,d^3x.$$

As $t\to\infty$, the second term vanishes by (i) above and the third one also vanishes by the Schwarz inequality.

Therefore,

$$I_C=\lim_{t\to\infty}\int_C|\Phi_t^C(x)|^2\,d^3x=\lim_{t\to\infty}\int_C|\Phi_t(x)|^2\,d^3x$$

by (ii). That is,

$$\lim_{t\to\infty}\int_C|\Phi_t(x)|^2\,d^3x=\int_{C'}|\hat\Phi(k)|^2\,d^3k.\qquad\text{Q.E.D.}$$

Now we prove a similar property for fully developing states. Note that the wave operators $\Omega_\pm$ exist and are complete under the assumption $V\in L^1(\mathbf{R}^3)\cap L^2(\mathbf{R}^3)$ (see Ref. 5). We follow Ref. 3. Given a cone C, define two projection operators F_C and G_C in $L^2(\mathbf{R}^3)$ by

$$(F_Cg)(x)=\chi_C(x)g(x),$$
$$(G_Cg)^\wedge(k)=\chi_C(\nabla\lambda(k))\hat g(k).$$

Then we know, by property 1,

$$\lim_{t\to\infty}\|F_Ce^{-itH_0}\Phi\|^2=\|G_C\Phi\|^2,\quad\text{for any}\quad\Phi\in L^2(\mathbf{R}^3).\tag{2.6}$$

Property 2: Let $g\in L^2(\mathbf{R}^3)$ with $\|g\|=1$. Then,

$$\lim_{t\to\infty}\|F_Ce^{-itH}\Omega_-g\|^2=\|G_CSg\|^2,$$

i.e., the probability that the scattered state (or state of absolute continuity) will be found in the cone C as $t\to\infty$ is given by

$$\int\chi_C(\nabla\lambda(k))|(Sg)^\wedge(k)|^2\,d^3k.$$

Here S is the S-operator $\Omega_+^*\,\Omega_-$.

Proof: By the definition of the wave operator Ω_+,

$$\lim_{t\to\infty}e^{-itH}e^{-itH_0}\Omega_+^*\,\Omega_-g=\Omega_+\Omega_+^*\,\Omega_-g,$$

so

$$\lim_{t\to\infty}\|e^{-itH_0}Sg-e^{-itH}\Omega_-g\|=0$$

since $S=\Omega_+^*\,\Omega_-$ and $\Omega_+\Omega_+^*\;=E_{ac}(H)$. In particular,

$$\lim_{t\to\infty}|\;\|F_Ce^{-itH_0}Sg\|-\|F_Ce^{-itH}\Omega_-g\|\;|$$
$$\leqslant\lim_{t\to\infty}\|e^{-itH_0}Sg-e^{-itH}\Omega_-g\|,\ \text{since}\ \|F_C\|\leqslant1$$
$$=\lim_{t\to\infty}\|e^{-itH}e^{-itH_0}Sg-\Omega_-g\|=\|\Omega_+Sg-\Omega_-g\|$$
$$=\|\Omega_+\Omega_+^*\,\Omega_-g-\Omega_-g\|=\|\Omega_-g-\Omega_-g\|=0.$$

Thus

$$\lim_{t\to\infty}\|F_Ce^{-itH}\Omega_-g\|^2=\lim_{t\to\infty}\|F_Ce^{-itH_0}Sg\|^2=\|G_CSg\|^2,$$

by Eq. (2.6). Q.E.D.

ACKNOWLEDGMENTS

This paper is based on the author's thesis at Indiana University. He is grateful to Professor R. G. Newton for suggesting the problem, helpful guidance, and constant encouragement. He also thanks C. H. Kim for help in revising the manuscript.

He was supported in part by the Indiana University Dissertation Fellowship and by the National Science Foundation.

[1]J. Dollard, Commun. Math. Phys. **12**, 193 (1969).

[2]J. M. Jauch, R. Lavine, and R. G. Newton, Helv. Phys. Acta **45**, 325 (1972).

[3]W. Amrein, J. Jauch, and K. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, Reading, MA, 1977).

[4]R. G. Newton, *Scattering Theory of Waves and Particles* (Springer, New York, 1982), 2nd ed.

[5]S. H. Yoon, "On impurity scattering in a periodically stratified potential in three dimensions," submitted to J. Math. Phys.

[6]N. Bleistein and R. Handelsman, *Asymptotic Expansion of Integrals* (Holt, Rinehart and Winston, New York, 1975).

On spherically symmetric shear-free perfect fluid configurations (neutral and charged). III. Global view

Roberto A. Sussman[a)]
School of Mathematical Sciences, Queen Mary College, Mile End Road, London E1 4NS, England

(Received 6 August 1987; accepted for publication 25 November 1987)

The topology of the solutions derived in Part I [J. Math. Phys. **28**, 1118 (1987)] is discussed in detail using suitable topological embeddings. It is found that these solutions are homeomorphic to $S^3\times\mathbb{R}$, $\mathbb{R}^4$, or $S^2\times\mathbb{R}^2$. Singularities and boundaries in these manifolds are examined within a global framework. One of these boundaries (mentioned but not examined in Part II [J. Math. Phys. **29**, 945 (1988)]) is regular (though unphysical), and is associated with an "asymptotically de Sitter" behavior characterized by an exponential form of the Hubble scale factor. Solutions with $S^2\times\mathbb{R}^2$ topology lack a center of symmetry [fixed point of SO(3)] and present a null boundary at an infinite affine parameter distance along hypersurfaces orthogonal to the four-velocity. This boundary, which in some cases is singular, can be identified as a null $\mathscr{I}$ surface arising as an asymptotical null limit of timelike hypersurfaces. Solutions with this topology, matched to a Schwarzschild or Reissner–Nordstrøm region, describe collapsing fluid spheres whose "surface" (as seen by observers in the vacuum region) has finite proper radius, but whose "interior" is a fluid region of cosmological proportions. In the case when the null boundary of the fluid region is singular, it behaves as a sort of "white hole." Uniform density solutions which are not conformally flat are all homeomorphic to $S^2\times\mathbb{R}^2$. Conformally flat solutions are also examined in detail. Their global structure has common features with those of FRW and de Sitter solutions. The static limits of all nonstatic solutions are discussed. In particular, under suitable parameter restrictions, some of these static solutions, together with the nonstatic conformally flat subclass, are the less physically objectionable of all solutions. Hence, it is suggested that kinetic theory models could be applied to them. Possible cosmological applications are discussed. The global structure of Wyman's and McVittie's solutions is examined in the Appendices.

I. INTRODUCTION

In most papers dealing with particular cases of ChKQ solutions (category "b" in the Introduction of Part I, see Ref. 1), these solutions are used as models of bounded collapsing spheres matched to Reissner–Nordstrøm (or Schwarzschild) space-time (see Sec. XI of Part II). Besides previous work by Cook[2] and Krasinski[3] on conformally flat solutions, by Mashhoon and Partovi[4] on charged configurations, and by these latter authors[5] and Collins[6] on the Wyman solution,[7] the global structure of ChKQ solutions (in the general case when they describe unbounded configurations) has been barely examined. Hence, by filling this gap in the literature, this paper aims to continue the study of this large class of solutions initiated in the preceding paper (Part II, see Ref. 1). As in Part II, the approach is extensive rather than intensive, that is, looking at general properties common to large subclasses instead of studying individual solutions in great detail. Subclasses of ChKQ solutions skipped in Part II, such as solutions with two time-dependent parameters (t-parameters, see Sec. V of Part II), uniform density, and conformally flat and static solutions, will be examined here. A brief description of the contents of each section is given below.

In Sec. II the topology (i.e., homeomorphic class of equivalence) of the hypersurfaces of constant coordinate time (i.e., the surfaces Σ_T), everywhere orthogonal to the four-velocity field, is studied. Knowing the topology of these three-surfaces, together with the information on the conformal structure of the inextendible boundaries and completeness of causal curves in these manifolds, allows one to deduce the topology of the latter. For this purpose, and remaining at the level of homeomorphic invariance, it is useful to perform suitable topological (but not isometric) embeddings of the hypersurfaces Σ_T and the space-time manifold $\mathscr{M}$ (as foliated by the Σ_T) into Euclidean spaces of one extra dimension. Such embeddings, which are introduced and discussed qualitatively in Secs. II and III, show how the topology of the space-time manifold depends on the number of regular centers, or world lines of fixed points of SO(3) (see Sec. VIII of Part II), which each solution admits. Solutions admitting two regular centers are homeomorphic to $S^3\times\mathbb{R}$, while those solutions admitting one or zero regular centers are homeomorphic to $\mathbb{R}^4$ and $S^2\times\mathbb{R}^2$, respectively. In some cases, these features are readily found from the coordinate representation, but other cases require a closer examination in terms of invariant quantities such as the affine parameter of geodesics along the Σ_T and the proper radius of the orbits of SO(3).

Section IV examines a boundary marked by finite coordinate values which was mentioned (but not studied) in Part II [$\Pi=0$, Eq. II(50)]. [Equation II(50) denotes Eq. (50) of Part II. All reference to equations of Parts I and II will be made in this form.] As comoving observers approach this

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

 0022-2488/88/051177-35$02.50

boundary, their proper time τ diverges, and so timelike geodesics (whose affine parameter is necessarily longer than τ) are complete. Since the Hubble scale factor H diverges as $\Pi \to 0$ taking an asymptotically exponential form in terms of τ, the boundary $\Pi = 0$ can be associated with a sort of "asymptotically de Sitter" behavior of comoving observers in their infinite past/future (i.e., a sort of "inflationary phase"). As R, the radius of the orbits of SO(3), also diary can be identified with a (future/past) null infinity surface $\mathscr{I}_{\pm}$ ("scri" plus/minus), which in some cases will be spacelike, like the $\mathscr{I}$ surface of de Sitter space-time,[8] though, depending on the parameters of the solutions, $\Pi = 0$ could be timelike, null, or without a globally defined conformal structure. In the representation of the topological embeddings of Sec. III, the space-time manifolds look like hyperboloidal shapes near $\Pi = 0$. Hence the change of topology in the Σ_T can be associated to the fact that these spacelike slices "stretch" all the way to $\Pi = 0$ in the infinite past/future of comoving observers. As reported by Cook[2] and Krasinski,[3] this effect occurs also in conformally flat solutions which are examined in Sec. IX. The Wyman solution[5-7] (see Appendix A) also exhibits this "asymptotically de Sitter" behavior.

Section V examines ChKQ solutions in which the locus $r = 0$ does not mark the world line of a regular center. It is found that such a locus marks in these solutions a null boundary at infinite affine parameter distance along geodesics in the slices Σ_T. This boundary, which can be identified as a null $\mathscr{I}$ surface, is in some cases singular. Some of these solutions have a regular center at $r = \pi$, or at $r = \infty$, and so are homeomorphic to $\mathbb{R}^4$. However, those solutions lacking a regular center have slices Σ_T with a "wormhole" $S^2 \times \mathbb{R}$ topology and the space-time manifold is then homeomorphic to $S^2 \times \mathbb{R}^2$. The well-known McVittie solution,[9] which is examined in Appendix B, belongs to the subclass of solutions discussed in this section.

The collapse of fluid spheres formed when a solution without a regular center is matched to a Schwarzschild or Reissner–Nordstrøm space-time is examined in Sec. VI. From the point of view of the evolution of the "surface" of the sphere in the vacuum region, the collapse picture is similar to that discussed in Sec. XI of Part II, with the surface of the sphere terminating its evolution in the "finite-density" (FD) singularity [solutions of type (ii)], or collapsing into a null "localized" (L) singularity [solutions of type (iv)]. However, the lack of a center in the fluid region means that the latter does not describe a stellar model and so, cannot be properly called an "interior" region. In fact the "inside" of this fluid sphere is a space-time region of cosmological dimensions, with its own $\mathscr{I}$ surface at $r = 0$. Hence in type (iv) solutions, one has the realization of the "science fiction" idea of "another universe" inside of a black hole, as observers inside of the collapsing sphere not only avoid the L singularity but have complete world lines and receive photons from infinity (the null $\mathscr{I}$). In the cases in which the null $\mathscr{I}$ is singular, light rays from this boundary can be received in the Schwarzschild or Reissner–Nordstrøm region, as if it were as a sort of white hole. These unusual features have been overlooked by authors, such as McVittie,[10] Knutsen,[11] Mashhoon and Partovi[12] (their Sec. VI), and Glass,[13] who have investigated the collapse of this type of spheres.

Section VII deals with ChKQ solutions with $\Psi_{(2)} \neq 0$ (i.e., not conformally flat) having two time-dependent parameters. It is shown that the extra t-parameter leads to a metric coefficient g_{tt} and a Raychaudhuri equation involving an elliptic integral of the second kind. Since a detailed study of the properties of these solutions would require numerical work or series approximations of these integrals, it is not attempted. Instead, it is argued, using a simple particular example, that some features of the singularity structure of these solutions might change in comparison with the case in which L is constant (W-type solutions, see Sec. IV of Part II). However, the topology of the solutions does not seem to be affected by the existence of the extra t-parameter.

Section VIII examines uniform density solutions with $\Psi_{(2)} \neq 0$ (UD solutions), all of which are particular cases of the solutions discussed in Sec. V, and are homeomorphic to $S^2 \times \mathbb{R}^2$. The case in which these solutions describe a neutral fluid in the presence of a source-free electric field is examined. Since this electric field is orthogonal to u^a, it is suggested that the wormhole topology of the slices Σ_T prevents the field lines converging into "sinks" or "sources," and so the neutral fluid behaves as a dielectric material with uniform polarization. Particular cases of UD solutions with such an electric field were examined in Sec. VI of Mashhoon and Partovi,[12] however, these authors did not suggest any explanation for the apparent lack of electric field sources. Neutral UD solutions have been discussed by Glass[13] and by Eisenstaedt,[14] but they ignored their singularity structure and global properties.

Conformally flat solutions (CF solutions) are studied in detail in Sec. IX, complementing the work of Cook[2] and Krasinski.[3] Depending on their parameters, these solutions might present the FD singularity or a standard big-bang qualitatively similar to that of FRW solutions. The occurrence of the latter singularity was overlooked by these authors. The global structure of CF solutions is examined in connection with the topology of the slices Σ_T, and with the help of the topological embeddings of Secs. II and III. These solutions share common global features with de Sitter or FRW solutions, and in particular, those solutions in which the boundary $\Pi = 0$ and the FD singularity do not arise have global features similar to a closed (i.e., $S^3 \times \mathbb{R}$) FRW cosmology. The latter case constitutes those ChKQ solutions which are less physically objectionable from a local and global point of view.

The static limits of all nonstatic ChKQ solutions are examined in Sec. X. In particular, it is found that the static limit of UD solutions is the Reissner–Nordstrøm (or Schwarzschild) solution with the parameter L as a cosmological constant, while CF solutions reduce to de Sitter solution or to the "interior" Schwarzschild solution. Other nonstatic ChKQ solutions reduce to well-defined static solutions, some of which are already known (Buchdahl's relativistic generalization of polytropic spheres of index 5, see Refs. 12, 15, and 16). In Sec. XI, the possibility of a kinetic theory approach is suggested for some ChKQ solutions in connection with the discussion in Sec. VI of Part II.

It is argued that kinetic theory configurations describing gasses with "detailed balance" collisions could be modeled on some static solutions (for example, Buchdahl's solution mentioned above). On the other hand, these static solutions and suitable nonstatic CF solutions could be used as models of collisionless gas mixtures, following previous work on this subject.[17–20]

In Sec. XII, the applicability of ChKQ solutions as cosmological models is discussed. In particular, it is argued that cosmological configurations proposed by Eisenstaedt[14] by matching CF and UD solutions are theoretically interesting but physically unacceptable from a local and global point of view. Since most ChKQ solutions either have privileged observers (i.e., observers comoving along a regular center) or have unphysical local and global features, it is suggested that the most suitable cosmological application would be to use the above-mentioned kinetic theory configurations as "Swiss cheese" models of local inhomogeneities in a FRW background.

The global structure of the Wyman solution[7] is examined in detail in Appendix A, as a continuation of Appendix A of Part II, and complementing the work of Collins[6] and of Mashhoon and Partovi[5] on this solution. Appendix B shows the global structure of McVittie's pioneering solution,[9] which has been suggested as describing a model of a "point particle" immersed in a cosmological fluid.[21–23] It is found that the global features of this solution are inconsistent with such a model. In Appendix C it is proven that different values of the parameter k in Eqs. II(2) correspond to the same solution if $\rho = \rho(T)$ (UD and CF solutions).

II. THE SURFACES Σ_T

As mentioned in Sec. II of Part II, the time coordinate in the metric II(1) [in whatever form II(31), II(32), or II(33)] denotes spacelike hypersurfaces, the surfaces Σ_t, invariantly characterized by being everywhere orthogonal to the four-velocity field u^α. Since these hypersurfaces are achronal sets, so that the time coordinate labeling them is a global time function, ChKQ solutions (and in general, SSSF solutions) are stably causal space-times.[24,25] In particular, the choice given by II(33), having the metric coefficient H fully determined as a function of the coordinates (T,r), is better suited to examine these solutions globally, and so unless stated otherwise, it will be adopted henceforth.

Let $\mathscr{M}$ denote generically any given ChKQ solution [a spherically symmetric space-time manifold with metric II(1)], the set of surfaces Σ_T, denoted as $\{\Sigma_T\}$, constitutes a foliation of $\mathscr{M}$. Hence each $\Sigma_T (T = T_0)$ is a three-dimensional (3-D) Riemannian submanifold of $\mathscr{M}$, "without edge," characterized by the induced metric

$$d\sigma^2 = H^2(T_0,r)[dr^2 + f^2(r)(d\theta^2 + \sin^2\theta\, d\phi^2)], \qquad (1)$$

where H is any of the forms given in Part I with $T = T_0$ fixed and $f(r)$ is given by the three choices of Eqs. II(2):

$$f(r) = \begin{cases} r, & k = 0, \\ \sin r, & k = 1, \\ \sinh r, & k = -1, \end{cases}$$

so that the surfaces Σ_T are manifestly conformal to three-surfaces of constant curvature. The latter type of three-surfaces are characterized by an induced metric like (1) with conformal factor $H = H_0 = \text{const}$ and 3-D Ricci tensor and scalar given by

$$^{(3)}\mathscr{R}_{ij} = \delta_{ij}\, {}^{(3)}\mathscr{R}, \qquad (2a)$$

$$^{(3)}\mathscr{R} = k/H_0, \qquad (2b)$$

where $^{(3)}\mathscr{R}_{ij}$ and $^{(3)}\mathscr{R}$ are not the four-dimensional Ricci tensor and scalar, $\mathscr{R}_{\alpha\beta}$ and $\mathscr{R}$, restricted to hypersurfaces of constant time, but the Ricci tensor and scalar computed with the three-metric g_{ij} [(1) with $H = H_0$]. In three-surfaces of constant curvature the values $k = 0, \pm 1$ mark whether this curvature is zero, negative, or positive, and so whether these hypersurfaces are isometric to the three-sphere S^3 ($k = 1$), to Euclidean space $\mathbb{R}^3$ ($k = 0$), or to the 3-D "pseudosphere" $\mathbb{H}^3$ ($k = -1$).

However, the comparison between Σ_T [in which $H = H(r)$] and three-surfaces of constant curvature might be misleading, as two conformally related manifolds might have very different local and global properties. For example, the form of the Ricci tensor and scalar given by Eqs. (2) is associated with well-known group invariance properties of three-surfaces of constant curvature. Such properties characterize globally these three-surfaces, but do not hold in general for surfaces Σ_T of ChKQ solutions (even in the case of uniform density solutions where $^{(3)}\mathscr{R}$ is constant, see Sec. VIII). For the surfaces Σ_T with $H = H(r)$, the Ricci scalar is in general the function of r given by II(18) [with $\Theta = \Theta(T_0)$ const], and thus might change sign locally [if $24\pi\rho(T_0,r) = \Theta^2(T_0)$] without conveying any information on the global structure of the slices Σ_T. Therefore, other intrinsic invariant quantities, besides $^{(3)}\mathscr{R}$, must be considered.

Being spherically symmetric submanifolds, the surfaces Σ_T can be described globally as warped products of the form $S^2 X_R \mathscr{C}$ (i.e., three-surfaces of revolution[26]), with $\mathscr{C}$ an open, semi-open or closed subset of $\mathbb{R}$, depending on whether Σ_T contains zero, one, or two centers [fixed points of SO(3)], respectively. The fibers of $S^2 X_R \mathscr{C}$ are then two-spheres [orbits of SO(3)], labeled by the coordinate r, whose proper area $4\pi R^2(r)$ and proper radius $R = fH$ are given in terms of the warping function R, which vanishes at a center. The leaves $\mathscr{C}$ are the set of "radial" curves with angular coordinates θ and ϕ fixed. Since the coordinate r parametrizing $\mathscr{C}$ has no geometric meaning besides providing a label of the orbits of SO(3), it is necessary to define an invariant measure of spacelike separation between the fibers S^2. Such an invariant measure can be the proper "arc length" of the leaves of $S^2 X_R \mathscr{C}$. Since the leaves in a warped product are totally geodesic submanifolds,[26] the above-mentioned radial curves are geodesics of Σ_T (though not geodesics of $\mathscr{M}$), and their proper length is an affine parameter length along Σ_T defined as a smooth (at least C^1) function $\zeta: \mathscr{C} \to \mathbb{R}$ by

$$\zeta(r) \equiv \int_{r_0}^{r} H(T_0,\bar{r})\, d\bar{r} = \int_{r_0}^{r} h(\bar{r})\, \frac{\Xi(T_0,\bar{r})}{\Pi(T_0,\bar{r})}\, d\bar{r}, \qquad (3)$$

where the generic form of H of II(43) has been used. Knowing the behavior of the invariants R and ζ along Σ_T is sufficient information to deduce its homeomorphic class of equivalence (i.e., its topology). However, it is important to know also if $\mathscr{M}$ is regular (in the sense of Sec. VIII of Part II) along a given Σ_T. Since the singular boundaries $H=0$ and $Q=0$ (or other boundaries, such as $\Pi=0$, see Sec. IV) do not coincide (in general) with any Σ_T (see Table I of Part II), some of the Σ_T in any given $\mathscr{M}$ do not "reach" these boundaries and so extend for all the regular range of r. However, there are always Σ_T in $\mathscr{M}$ "reaching" either one (or two) of the boundaries, and so will not cover all the regular range of r.

A usual way to "visualize" three-surfaces like Σ_T is to embed them in $\mathbb{R}^4$, preferentially performing an isometric embedding by demanding that the induced metric of Σ_T in $\mathbb{R}^4$ coincides with (1). However, an isometric embedding of Σ_T in $\mathbb{R}^4$ is not always possible, and if it is, it requires using R as radial coordinate in (1) (see Sec. 23.8 in Ref. 20), and so $R(T,r)$ must be inverted as $r(R,T)$ which, for the forms of H obtained in Part I, cannot be done explicitly in general. Though, in order to appreciate the structure of Σ_T at the level of topology, the geometric meaning of R can be used to perform a topological (though not isometric) embedding in which $R(r)$ generates a surface of revolution in $\mathbb{R}^4$. Such an embedding is given by the map

$$\mathscr{Y}:\ \Sigma_T \to \mathbb{R}^4,\qquad (r,\theta,\phi)\to(r,R(r),\theta,\phi) \tag{4}$$

where $\mathbb{R}^4$ has the metric element

$$ds^2 = dr^2 + dR^2 + R^2(d\theta^2 + \sin^2\theta\, d\phi^2),$$

so that the surface of revolution $\mathscr{Y}(\Sigma_T)$ is homeomorphic to Σ_T. A 2-D representation of Σ_T under the embedding (4) will be displayed further ahead for various forms of Σ_T. It will be assumed in this and the following section that $\{\Sigma_T\}$ foliate M- or W-type solutions (see Sec. IV of Part II) for which conditions II(45) hold at $r=0$, $|\mathrm{II}(T,r)|>0$, and $\rho'\neq 0$. The cases excluded here will be examined in Secs. IV–VIII, the cases $k=1$ and $k=0,-1$ in Eq. (1) will be treated separately.

A. Case $k=1$

In this case the Σ_T not reaching the boundaries extend along the range $0\leqslant r\leqslant\pi$ [$T=T_3$ in Figs. 1(a) and 1(b)]. Since $R(T_3,0)=R(T_3,\pi)=0$ and $R(T_3,r)>0$ for $0<r<\pi$, and so $\mathscr{C}=[0,\pi]$, Σ_T contains two regular centers [fixed points of SO(3)]. Thus Σ_T is in this case homeomorphic to S^3 and it can be embedded in $\mathbb{R}^4$ through the map (4) as a sort of "deformed" spheroidal shape in which $r=0$ and $r=\pi$ mark the two "antipodal" centers. A 2-D representation of $\mathscr{Y}(\Sigma_T)$ is shown in Fig. 2(a). Since Σ_T is compact, it is also geodesically complete.[26]

If Σ_T reaches the FD singularity $Q=0$, at say, $r=r_1$, Σ_T extends along the range $0\leqslant r<r_1$ [see $T=T_1$ in Fig. 1(a)]. Thus one has $R(T_1,0)=0$, $R(T_1,r)>0$ for $0<r<r_1$, and r_1 marks a two-sphere of radius $R(T_1,r_1)=A/(f_0h_0)$ [A is a root of Q in I(24a)] at which $\mathscr{M}$ is singular. Hence Σ_T is in this case homeomorphic to $\mathbb{R}^3$, and it can be visualized as a sort of "opened" or "cracked" surface of revolution in the embedding diagram of Fig. 2(b). If Σ_T reaches the AD big bang $H=0$ at r_1, extending along $0\leqslant r<r_1$ [see $T=T_1$ in Fig. 1(b)], then one has $R(T_1,r)>0$ for $0<r<r_1$ and $R(T_1,0)=R(T_1,r_1)=0$, but $r=r_1$ marks a singular point (curvature scalars of $\mathscr{M}$ diverge even if causal curves are complete) and so it is not contained in Σ_T. Hence Σ_T is also homeomorphic to $\mathbb{R}^3$, looking in the embedding diagram of Fig. 2(c) like a sort of a surface of revolution "punctured" at $r=r_1$. A third possibility arises if Σ_T reaches both singular boundaries, so that the regular range of r is $r_2<r<r_1$ with $r_2>0$ and $r_1<\pi$ [see Fig. 1(c)]. In this case Σ_T is homeomorphic to $S^2\times\mathbb{R}$ with the embedding diagram shown in Fig. 2(d). Since H is finite as $r\to r_2$ and/or $r\to r_1$ whenever Σ_T reaches a singular boundary, ζ evaluated from (3) is finite and so Σ_T is geodesically incomplete in the cases illustrated by Figs. 2(b)–2(d).

B. Solutions with $k=0,-1$

In this case one has $R(T,0)=0$, and if Σ_T does not reach a boundary, the range of r is $0\leqslant r<\infty$ [$T=T_2$ in Figs. 3(a) and 3(b)]. However, $R(T_2,r)$ might vanish as $r\to\infty$, and infinite values of r might be at a finite affine parameter distance. Hence, for $R(T_2,r)>0$ ($0<r<\infty$), one has the following four possibilities as $r\to\infty$:

$$R\to 0,\quad \zeta \text{ finite}, \tag{5a}$$

$$R \text{ finite},\quad \zeta \text{ finite}, \tag{5b}$$

$$R \text{ finite},\quad \zeta\to\infty, \tag{5c}$$

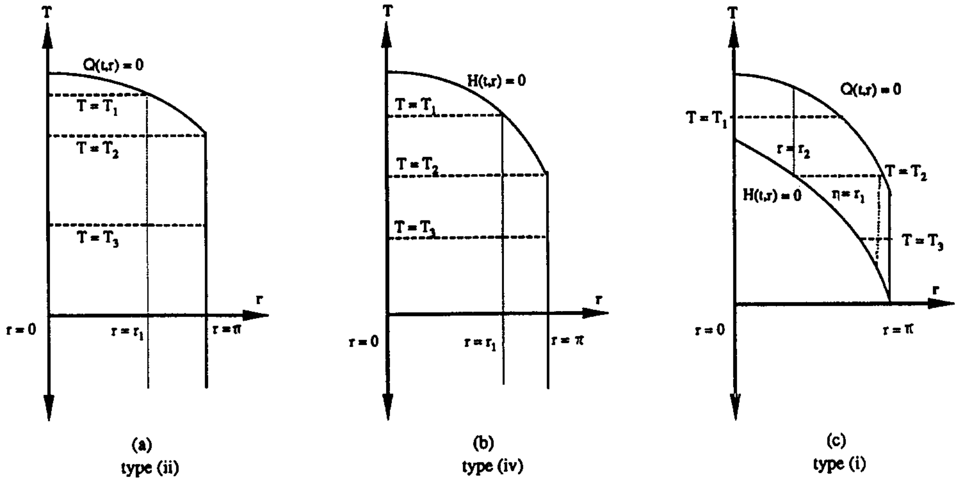


FIG. 1. Domain of regularity of solutions with $k=1$. Various surfaces Σ_T and Σ_r are represented as dotted horizontal and vertical lines. Solutions with $k=0,-1$ belonging to the case given by Eq. 5(a) have similar domains of regularity, with $r=\pi$ replaced by $r=\infty$.

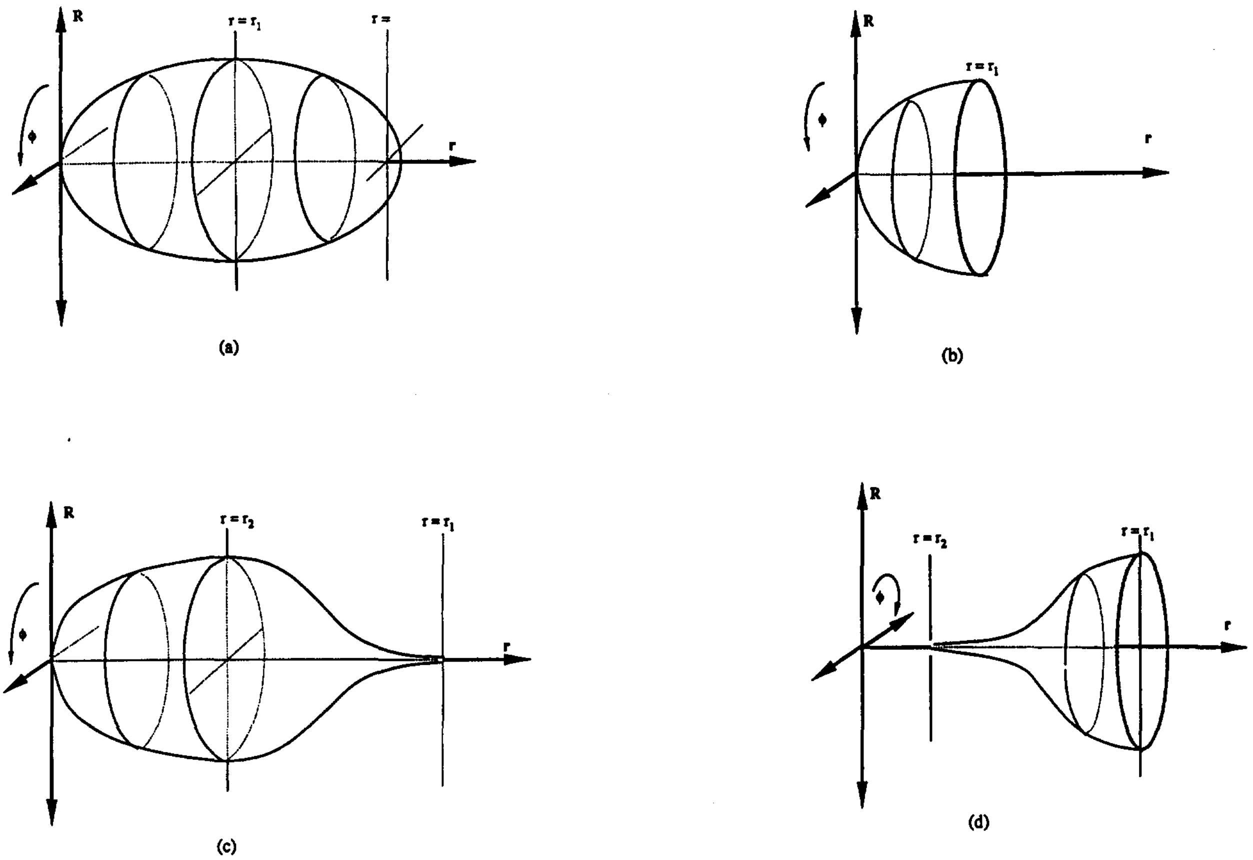


FIG. 2. Representation of the embedding of surfaces Σ_T in the case $k = 1$. A 2-D representation of $\mathscr{Y}(\Sigma_r)$, as surfaces of revolution in $\mathbb{R}^3$ with profile $R(r)$, can be obtained by setting $\theta = \pi/2$ in Eq. (4). (a) represents a complete surface Σ_T homeomorphic to S^3 [$T = T_3$ in Figs. 1(a) and 1(b)]. (b) and (c) depict the incomplete surfaces $T = T_1$, homeomorphic to $\mathbb{R}^4$ of Figs. 1(a) and 1(b). (d) corresponds to $T = T_2$, homeomorphic to $S^2 \times \mathbb{R}$, in Fig. 1(c).

$$R \to \infty, \quad \zeta \to \infty. \tag{5d}$$

In the case (5a), $T = T_2$ is homeomorphic to S^3 with $r = \infty$ denoting an "antipodal" center [fixed point of SO(3)] analogous to $r = \pi$ in the case $k = 1$ [i.e., $T = T_3$ in Figs. 1(a) and 1(b)]. Therefore, the same arguments as in the case $k = 1$ apply (see Figs. 1 and 2), and there is always a suitable relabeling $\bar{r} = \bar{r}(\mathrm{r})$ so that $\bar{r}(\infty)$ is finite and $\mathscr{C} = [0, \bar{r}(\infty)]$. In the case (5b), $r = \infty$ marks the end of a coordinate patch and so the topology of $T = T_2$ cannot be known unless the coordinate r is analytically extended [see Fig. 4(a)]. In cases (5c) and (5d), $T = T_2$ is geodesically complete and since $\mathscr{C} = [0, \infty)$, it is clearly homeomorphic to $\mathbb{R}^3$. The embedding diagram for these cases is shown in Figs. 4(b) and 4(c).

In cases (5c) and (5d) above, surfaces Σ_T reaching the singular boundaries $Q = 0$ or $H = 0$, or both, might be homeomorphic to $\mathbb{R}^4$ [$T = T_2$ in Figs. 3(a) and 3(b)] or to $S^2 \times \mathbb{R}$ [$T = T_1$ in Figs. 3(a) and 3(b), and $T = T_2$ in Fig. 3(c)]. However, these surfaces [see Figs. 4(d)–4(f)] are geodesically incomplete, since $H(r_0)$ [and thus $\zeta(r_0)$] is finite for $r = r_0$ labeling the singular limit of the range of r along these Σ_T.

In W-type solutions, and in general if $L \neq 0$, the functions Π and Ξ in Eq. (3) are of the same order in r as $r \to \infty$ [see Eqs. I(28), I(30), I(32), I(45), and I(47)]. Hence, as

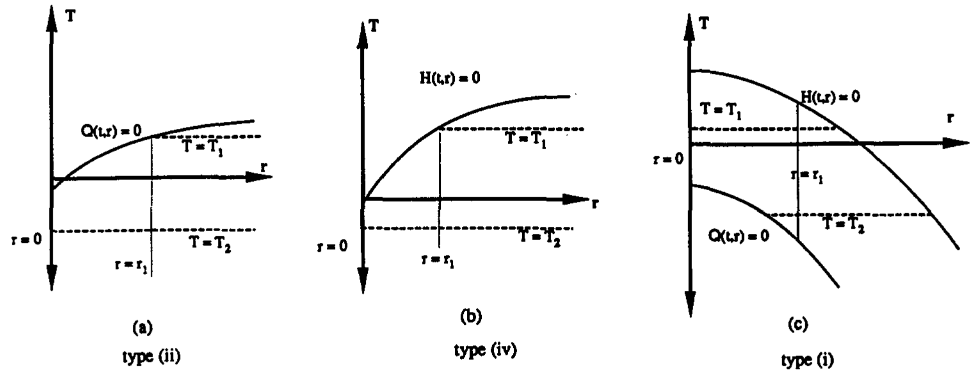


FIG. 3. Domain of regularity of solutions with $k = 0, -1$. This figure is the equivalent of Fig. 1 when r can take infinite values. As in Fig. 1, various surfaces Σ_T and Σ_r are displayed as dotted horizontal and vertical lines. However, if ζ (the affine parameter along the Σ_T) behaves as in Eq. (5a), the domain of regularity would be as in Fig. 1 with $r = \infty$ replaced by $r = \pi$.

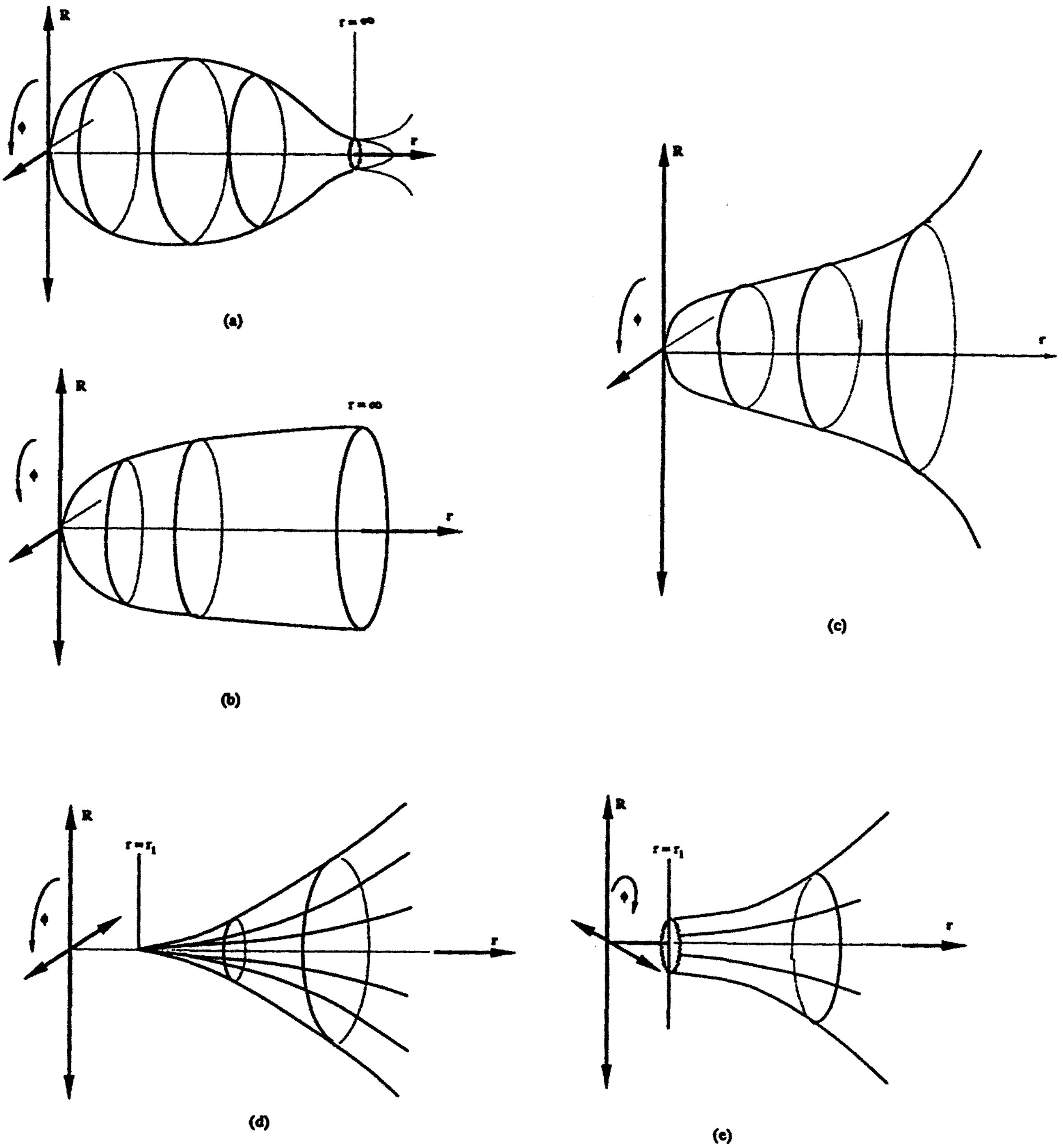


FIG. 4. Representation of the embedding of surfaces Σ_T in the case $k = 0, -1$. This figure provides a 2-D representation of $\mathscr{Y}(\Sigma_r)$ (as in Fig. 2) for various surfaces Σ_T shown in Fig. 3. (a) represents the case described by Eq. (5b), in which r must be extended in order to find out if Σ_T is homeomorphic to S^3 or $\mathbf{R}^3$. (b) and (c) correspond to the cases of Eqs. (5c) and (5d) describing complete surfaces Σ_T homeomorphic to $\mathbf{R}^3$ [$T = T_2$ in Figs. 3(a) and 3(b)]. (d) and (e) correspond to incomplete surfaces $T = T_1$, homeomorphic to $S^2 \times \mathbf{R}$, of Fig. 3. For the case described by Eq. (5a), the embedding diagrams of the Σ_T are as those of Fig. 2, with $r = \pi$ replaced by $r = \infty$.

$r \to \infty$, one has $\zeta \sim \int h\, dr$ and $R \sim fh$, and thus the topology of those Σ_T extending towards infinite values of r, as determined by the different cases (5a), (5c), and (5d), follows from the behavior as $r \to \infty$ of fh and h given by Eqs. I(17) and I(25). If $k = 0$ and h is given by I(25a) or I(25d), Σ_T is homeomorphic to S^3 as in case (5a); however, for other combinations of $k = 0, -1$ and h, one has either one of the cases (5b), (5c), or (5d) in which Σ_T is homeomorphic to $\mathbf{R}^3$ or the coordinate r must be analytically extended. For M-type solutions, in which the function Π is a constant along the Σ_T, it is difficult to grasp at first hand the behavior of R and ζ as $r\to\infty$, and so the integral in (3) must be explicitly evaluated.

III. THE GLOBAL VIEW

Simply connected spherically symmetric space-times are globally decomposable into simple topological products of the form $\mathbb{B}\times S^2$, where $\mathbb{B}$ is a 2-D Lorentzian surface orthogonal to the orbits of SO(3) (Refs. 27 and 28). Since space-time manifolds which are not simply connected have no physical interest,[24] it will be assumed that ChKQ solutions are either simply connected, or have simply connected universal coverings. As the coordinate patches (T,r) or (τ,r) [τ defined by Eq. II(12), θ and ϕ fixed] in Figs. 1 and 3 (and also those used in Part II) are diffeomorphic to $\mathbb{B}$, it should be possible to deduce the topology of $\mathscr{M}$ from the information contained in these coordinate patches together with the knowledge of the topology of the surfaces Σ_T. In particular, the change of topology of the Σ_T (entirely due to the presence of singular boundaries) can be ascribed to the specific way in which the latter slice the space-time manifold.

A useful global representation of the Σ_T was obtained in the previous section through the topological embedding (4) by using the geometric invariant meaning of the function R. A similar global representation of $\mathscr{M}$, which is useful at the level of topological structure, can be constructed by performing a topological (but, again, not isometric) embedding of the latter (thought of as the set $\{\Sigma_T\}$) into $\mathbb{M}^5$, a 5-D semi-Riemannian space with signature $(-,+,+,+,+)$. However, the time evolution of $\mathscr{M}$ should be conveyed into this embedding in an invariant manner, and since T is merely a coordinate label, the parametrization of $\mathbb{B}$ by the proper time τ of comoving observers is better suited. If τ is formally defined as the function τ: $\mathscr{M}\to\mathbb{R}$ given explicitly by Eq. II(53), together with the function R: $\mathscr{M}\to\mathbb{R}$ given as $R=fH$, a convenient way to appreciate $\mathscr{M}$ globally is provided by the following topological embedding:

$$\begin{aligned}\mathscr{E}:\ &\mathscr{M}\to\mathbb{M}^5\\ &(T,r,\theta,\phi)\to(\tau(T,r),R(T,r),r,\theta,\phi),\end{aligned}\tag{6}$$

where $\mathbb{M}^5$ has the metric element

$$d\eta^2_{(5)}=-d\tau^2+dr^2+dR^2+R^2(d\theta^2+\sin^2\theta\,d\phi^2),$$

so that $\mathscr{E}(\mathscr{M})$, as a 4-D hypersurface of $\mathbb{M}^5$, is homeomorphic to $\mathscr{M}$. The topology of ChKQ solutions follows by qualitatively examining the homeomorphic class of equivalence of $\mathscr{E}(\mathscr{M})$ through suitable 2-D representations of the latter. As in the former section, this section only considers M- and W-type solutions in which $|\Pi|>0$ and II(45) hold at $r=0$. The cases $k=1$ and $k=0,-1$ are considered separately.

A. Case $k=1$

In these solutions $\mathscr{M}$ is homeomorphic to the "hypercylinder" $S^3\times\mathbb{R}$, shown as various forms of truncated cone-like objects in the 2-D representations of $\mathscr{E}(\mathscr{M})$ given in Figs. 5 and 6. For solutions of type (ii) [see Fig. 1(a)], the Σ_T reaching the FD singularity $Q=0$ are incomplete and homeomorphic to $\mathbb{R}^3$ because they "tilt" in such a way as to terminate in this singular boundary [$T=T_1$ in Fig. 5(b)]. In solutions of type (iv) [see Fig. 1(b)], as shown in Fig. 5(c), the "hypercylinder" $S^3\times\mathbb{R}$ becomes an infinitely long "neck" due to the fact that those Σ_T reaching the AD big bang $H=0$ do so as $\tau\to\infty$, and thus appear strongly tilted in the representation of $\mathscr{E}(\mathscr{M})$. Since the "point at infinity"

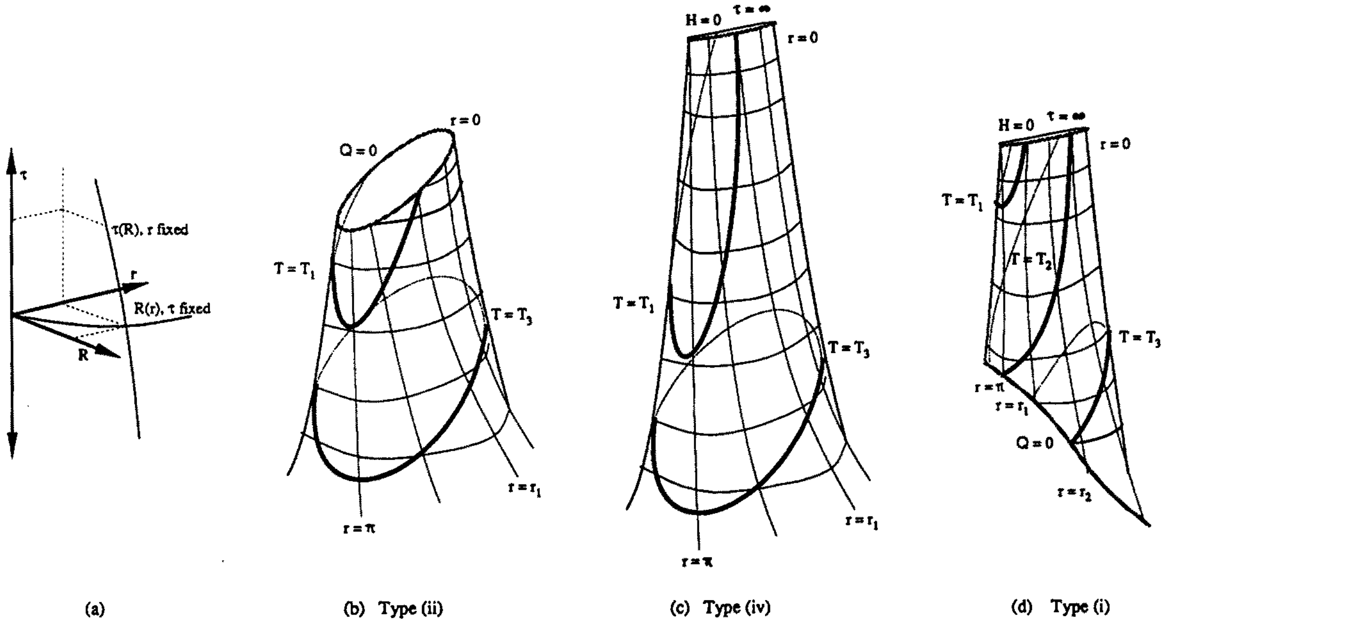


FIG. 5. Representation of the embedding of $\mathscr{M}$ for solutions with two regular centers. These 3-D representations of $\mathscr{E}(\mathscr{M})$ are obtained by setting $\theta=\pi/2$ and $\phi=0$ in Eq. (6). As shown in (a), vertical curves are the profile $\tau(R)$ for fixed r, while horizontal curves give the profile $R(r)$ for fixed τ. Additionally, for each curve $\tau(R)$ ($r=r_0$, $\theta=\pi/2$, $\phi=0$), one can associate the curve $\tau(R)$ corresponding to $r=r_0$, $\theta=\pi/2$, and $\phi=\pi$. These curves are plotted as "mirror images" of those with $\phi=0$, leading to the objects of (b)–(d). Various surfaces Σ_T are depicted as thick, tilted curves. The complete surfaces $T=T_3$, homeomorphic to S^3, of Figs. 1 and 2 appear as closed circles. Incomplete surfaces $T=T_1$ and $T=T_2$, homeomorphic to $\mathbb{R}^3$ and $S^2\times\mathbb{R}$ appear as open curves. In all cases, the space-time manifold $\mathscr{M}$ is homeomorphic to $S^3\times\mathbb{R}$.

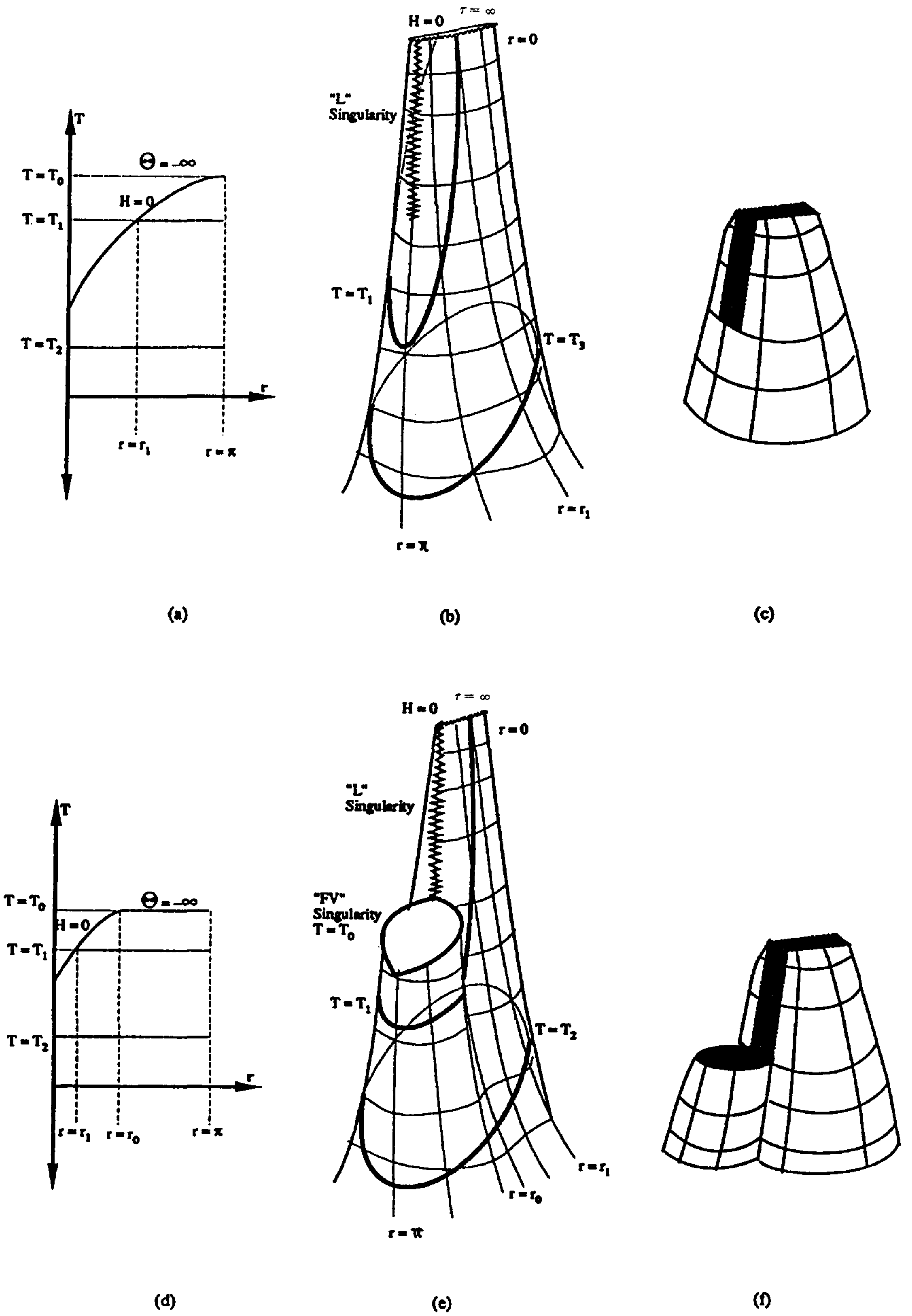


FIG. 6. Representation of the embedding of $\mathscr{M}$ for solutions with L and FV singularities. Consider a type (iv) solution whose domain of regularity is shown in (a) and (d). If $\Theta(T_0) \to -\infty$ along $r = \pi$, as shown in (a), the boundary $H = 0$ consists of a L singularity and an AD big bang, while if $\Theta(T_0) \to -\infty$ along $r_0 < r < \pi$, as in (d), one has a combination of L and FV singularities and AD big bang. A representation of $\mathscr{E}(\mathscr{M})$, analogous to those shown in Fig. 5, is displayed in (b) and (e) for both cases, with the L and FV singularities looking like a rigged line and a hole in $\mathscr{M}$. Notice how the surface $T = T_1$ bends upwards reaching the AD big bang at $\tau = \infty$. Both representations are homeomorphic to the objects with topology $S^1 \times \mathbf{R}$ shown in (c) and (f), indicating that the L and FV singularities do not change the $S^3 \times \mathbf{R}$ topology that $\mathscr{M}$ would have if these singularities were absent [see Fig. 5(c)]. If the solution has only one regular center (say at $r = \pi$), the embedding diagram would be similar to (b), except that the horizontal profile curves $R(r)$ would not close around the second center $r = 0$ [see Fig. 7(b)].

representing $H=0$ in Fig. 5(c) is not part of $\mathscr{M}$, then one has a similar situation to a cone in $\mathbb{R}^3$ in which the singular point at the top has been removed. For solutions of type (i) [see Fig. 1(c)], and as shown in the representation of Fig. 5(d), some of the Σ_T are incomplete and homeomorphic to $S^2\times\mathbb{R}$ [as in Fig. 2(d)], starting in $Q=0$ and tilting all the way toward $H=0$ as $\tau\to\infty$.

In solutions presenting the AD big bang $H=0$, the representation of $\mathscr{E}(\mathscr{M})$ shows this timelike singular boundary as a singular "slit" located in the infinite top end of long and thin "neck" [see Figs. 5(c) and 5(d)]. It is interesting to notice that the presence of the other singularities discussed in Sec. X of Part II, as seen through the embedding (6) do not change the $S^3\times\mathbb{R}$ topology of $\mathscr{M}$. For instance, if a L singularity emerges at $r=\pi$, $T=T_0$ [see Fig. 6(a)], then the representation of $\mathscr{E}(\mathscr{M})$ given by Fig. 6(b) shows this null singularity as a singular "line" extending from $\tau(T_0)$ toward $\tau\to\infty$ at the AD big bang $H=0$, that is, as a sort of "scratch" across $\mathscr{E}(\mathscr{M})$. If a spacelike FV and null L singularities arise as Θ diverges at $T=T_0$ [see Fig. 6(d)], the situation is that depicted by Fig. 6(e), which shows the combination of these singularities as a "hole" and a scratch in $\mathscr{E}(\mathscr{M})$. In both cases M is homeomorphic to $S^3\times\mathbb{R}$ just as the "rigged" and "punched" cones of Figs. 6(c) and 6(f) are homeomorphic to $S^1\times\mathbb{R}$.

B. Case $k=0,-1$

In the cases given by Eq. (5a), $\mathscr{M}$ is homeomorphic to $S^3\times\mathbb{R}$ and the arguments above apply, with $r=\pi$ replacing $r=\infty$ as the antipode center to $r=0$. However, in cases given by Eqs. (5c) and (5d), $\mathscr{M}$ is homeomorphic to $\mathbb{R}^4$, shown in Fig. (7) as "opened" sheets connected through the center $r=0$. For solutions of type (ii) [see Fig. 3(a)], the FD singularity $Q=0$ looks in the representation of $\mathscr{E}(\mathscr{M})$ given by Fig. 7(a) like a "crack" in the opened sheets. The surfaces Σ_T not reaching $Q=0$, being complete and homeomorphic to $\mathbb{R}^3$, appear in this representation as curves opening towards $r\to\infty$ and closing up at $r=0$. The incomplete surfaces Σ_T, whose topology is $S^2\times\mathbb{R}$ or $\mathbb{R}^3$, depending on whether they close around $r=0$, appear as open or disconnected curves. For solutions of type (iv) [see Fig. 3(b)], those Σ_T which avoid $H=0$ behave in the embedding diagram of Fig. 7(b) in an analogous manner as similar Σ_T surfaces in the case of solutions of type (ii). However, those Σ_T reaching this singular boundary are strongly tilted bending all the way towards $\tau\to\infty$. For solutions of type (i) [see Fig. 3(c)], the representation of $\mathscr{E}(\mathscr{M})$ is shown in Fig. 7(c).

As in the case $k=1$, the AD big bang $H=0$ looks in the representations of $\mathscr{E}(\mathscr{M})$ as a singular slit; however, if $k=0,-1$ [except the case (5a)], this slit [see Figs. 7(b) and 7(c)] is infinite as it prevents the closing of those Σ_T reaching it. If a null L singularity arises, it would appear in the representation of $\mathscr{E}(\mathscr{M})$ in a similar way as in Fig. 6(b) as a sort of scratch extending all the way towards $H=0$ as $\tau\to\infty$. If a spacelike FV singularity arises together with the L singularity, then these features would appear in the representation of $\mathscr{E}(\mathscr{M})$ as a crack and a scratch, as in Fig. 6(e). However, in all these cases the topology of $\mathscr{M}$ would remain $\mathbb{R}^4$.

From the evaluation of the integral (3) and the embeddings (4) and (6), the topology of the M- and W-type solu-

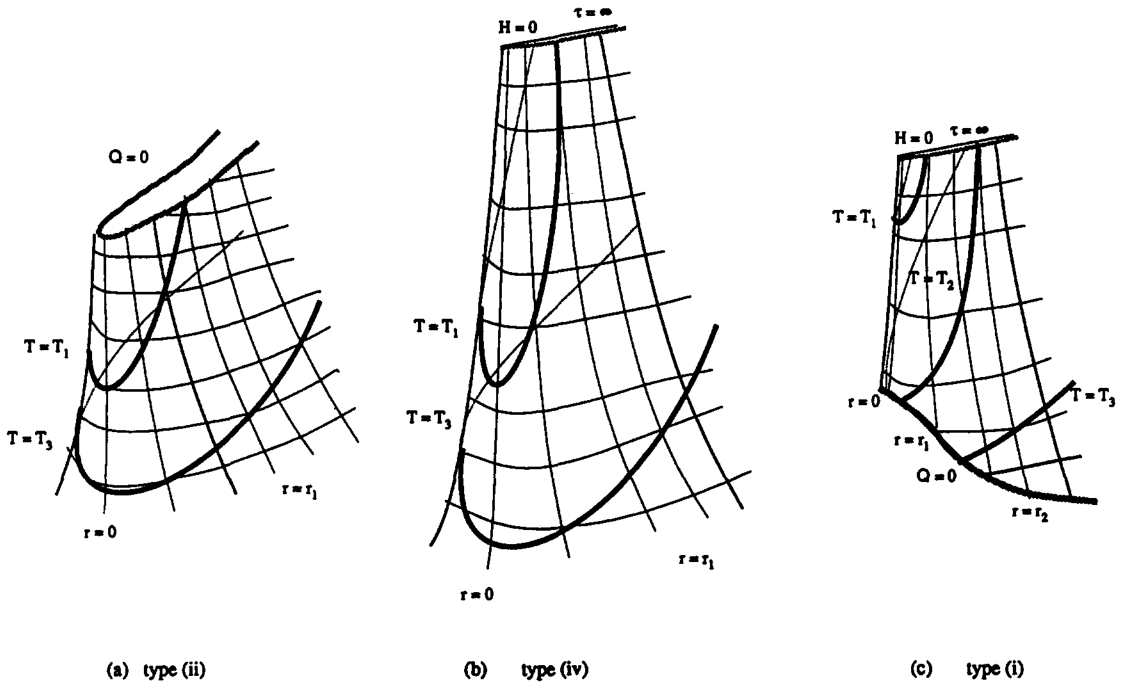


FIG. 7. Representation of the embedding of $\mathscr{M}$ for solutions having only one regular center at $r=0$. This representation of $\mathscr{E}(\mathscr{M})$ is equivalent to that of Fig. 5. As in the latter figure, vertical and horizontal profile curves denote surfaces Σ, and surfaces of constant τ. Various surfaces Σ_T of Figs. 3 and 4 are depicted as thick, tilted open curves. The surfaces $T=T_2$ are complete and homeomorphic to $\mathbb{R}^3$, while surfaces $T=T_1$ are incomplete and have topology $S^2\times\mathbb{R}$. In all cases, $\mathscr{M}$ is homeomorphic to $\mathbb{R}^4$.

TABLE I. Topology of M-type and W-type solutions. This table displays the topology of solutions classified in Tables III and IV in terms of the values of k [choice of $f(r)$ in Eq. II(2)] and of the form of the functions h and X defined by Eqs. I(23) and I(25). The fourth column from left to right is concerned with solutions discussed in Sec. V, hence, UD solutions (uniform density with $\Psi_{(2)} \neq 0$) correspond to the row $X2$ with $k = 0$.

Form of h and X in Eqs. I(25) and (11)	Value of k	Topology of solutions with a regular center at $r = 0$	Topology of solutions with a null boundary at $r = 0$	Behavior of W and Q in Eqs. (13)
$X1$ $a \neq 0$ $\Delta > 0$	$k = 0,1$	$c \neq 0$ $S^3 \times \mathbb{R}$	$c = 0$ [Eq. (11a)] $\mathbb{R}^4$	As in Eq. (13a), except NMcV($r2$)($X1$) and ChMcV($r2$)($X1$)
	$k = -1$	$\mathbb{R}^4$	$S^2 \times \mathbb{R}^2$	which correspond to Eq. (13c)
$X2$ $a = 0$ $b2 = \Delta$	$k = 1$	$c \neq 0$ $S^3 \times \mathbb{R}$	$c = 0$ [Eq. (11b)] $\mathbb{R}^4$	As in Eq. (13a) except NMcV($r2$)($X2$) and ChMcV($r2$)($X2$)
	$k = 0, -1$	$\mathbb{R}^4$	$S^2 \times \mathbb{R}^2$	which correspond to Eq. (13c)
$X4$ $a \neq 0$ $\Delta = 0$	$k = 0,1$	$b \neq 0$ $S^3 \times \mathbb{R}$	$b = 0$ [Eq. (11c)] $\mathbb{R}^4$	As in Eq. (13c) if $L = 0$. As in Eq. (13a) if $L \neq 0$
	$k = -1$	$\mathbb{R}^4$	$S^2 \times \mathbb{R}^2$	
$X5$ $a = b = 0$ $\Delta = 0$	$k = 1$	$S^3 \times \mathbb{R}$		
	$k = 0, -1$	$\mathbb{R}^4$		
$X3$ $a \neq 0$ $\Delta < 0$	$k = 1$	$S^3 \times \mathbb{R}$		
	$k = 0$	$S^3 \times \mathbb{R}$ (if $L \neq 0$) $\mathbb{R}^4$ (if $L = 0$)		
	$k = -1$	$\mathbb{R}^4$		

tions classified in Table I of Part II can be related to the value of k plus the form of the functions h and X in Eq. I(25). This relation is shown in Table I for the solutions classified in Tables III and VI of Part I.

C. Global hyperbolicity

As shown by the topological embedding (6), M- and W-type solutions with a regular center at $r = 0$ and in which $|\Pi| > 0$ holds are homeomorphic to $\Sigma_T \times \mathbb{R}$, where Σ_T is any Σ_T not reaching the boundaries of $\mathscr{M}$. Hence, these solutions seem to have a simple global structure, fully determined by the topology of the Σ_T reaching the existing (one or two) centers [fixed points SO(3)]. This situation (as far as the topology is concerned) is similar to the particular case when a space-time manifold $\mathscr{M}$ is globally hyperbolic, admitting a Cauchy hypersurface $\mathscr{C}$, in which case $\mathscr{M}$ is homeomorphic to the product $\mathscr{C} \times \mathbb{R}$ (see Ref. 25). However, since $\mathscr{M}$ is globally hyperbolic if and only if it has acausal (i.e., spacelike or null) boundaries,[29,30] then some ChKQ solutions, such as solutions of types (i) and (iv) presenting the timelike singular boundary $H = 0$, are definitely not globally hyperbolic. This is a consequence of the fact that global hyperbolicity implies stable causality, but the converse is false.

However, even for solutions of type (ii) and (iii), whose only boundary is (apparently) the spacelike FD singularity, global hyperbolicity will not hold if there are other (yet) undetected timelike boundaries, such as a timelike null infinity surface $\mathscr{I}$ ("scri") similar to that of the anti-de Sitter solution.[8] The occurrence of such a timelike $\mathscr{I}$ is possible, especially if $\mathscr{M}$ is homeomorphic to $\mathbb{R}^4$, though it will certainly be absent in solutions of type (iii) homeomorphic to $S^3 \times \mathbb{R}$ [see Fig. 1(c) of Part II], the latter being without doubt globally hyperbolic. Solutions of types (i), (ii), or (iv), homeomorphic to $S^3 \times \mathbb{R}$, might also be globally hyperbolic, especially if Θ vanishes or diverges at a given Σ_T so that the fluid bounces between a past and a future FD or FV singularities avoiding $H = 0$ (see Fig. 1 of Part II). These singularities can be subjected to the formalism of Wald and Yip[30] in order to verify if they admit a global parametrization by normal coordinates. However, solutions homeomorphic to $\mathbb{R}^4$ are likely to have $\mathscr{I}_\pm$ surfaces, and so the question of their global hyperbolicity requires further examination connected to verifying the conformal structure of these boundaries. This will not be attempted in the present paper.

IV. ASYMPTOTICALLY DE SITTER SOLUTIONS

From the form of the expressions of H derived in Part I, the boundary $\Pi = 0$ [Eq. II(50)] occurs in those W-type solutions ($L = \text{const} > 0$) and solutions with time-dependent L (see Secs. VII and IX) in which the function $\Pi(T,r)$ has real zeros. Some M-type solutions, such as NMcV($r2$)($X1,2$) and its charged version ChMcV($r2$)($X1,2$) present a similar boundary at $T = 0$ (see Appendix B and Table II). This regularity boundary

TABLE II. Solutions presenting the boundary $\Pi(T,r)=0$. This table complements Table I of Part II by giving the coordinate representation of the domain of regularity and boundaries for solutions presenting the boundary $\Pi=0$ (see Sec. IV). Parameters μ, ϵ, Δ and functions h, X, and u corresponding to each specific solution are those found in Tables III, VI, and VIII of Part I.

Solution	Type	Domain of regularity	$\Pi(T,r)=0$	$Q(T,r)=0$	$H(T,r)=0$
NMcV($r2$)(X1,2)	ii	$0<T<(\mu/2)u$	$T=0$	$T=(\mu/2)u$	
	iv	$0<T<(\lvert\mu\rvert/2)u$	$T=0$		$T=(\lvert\mu\rvert/2)u$
ChWy($r2$,$r2$)(X1,2)	ii	$-\frac{1}{u}<T<\frac{u-\Delta/2\epsilon^2}{(\Delta/2\epsilon^2)u-1}$, $T+u>0$	$T=-u$	$T=\frac{u-\Delta/2\epsilon^2}{(\Delta/2\epsilon^2)u-1}$	
ChWy($r2$,$r2$)(X3)	iv	$-u<T<u-1$	$T=-u$		$T=u-1$
ChMcV($r2$)(X1,2)	iv	$0<T<(\lvert\mu\rvert+\sqrt{\Delta\epsilon})u/2$	$T=0$		$T=(\lvert\mu\rvert+\sqrt{\Delta\epsilon})u/2$
ChWy($r2$)(X1,2)	ii	$\sqrt{u_0}<T+X<2\sqrt{u_0\beta}$ β as in Table I of Part II	$T=\sqrt{u_0}-X$	$T=2\sqrt{u_0\beta}-X$	
ChWy($r2$)(X4,5)	iv	$(\sqrt{2}-1)u<T<(\sqrt{3}-\sqrt{2})u$	$T=(\sqrt{2}-1)u$		$T=(\sqrt{3}-\sqrt{2})u$
ChWy($r2$)I(X1,2)	ii	$\sqrt{2}/(\sqrt{2}-1)<\cos V<1$ $V\equiv\sqrt{(\Delta/2)}[T+X]$	$V=\pi$	$V=\cos^{-1}[\sqrt{2}/(\sqrt{2}-1)]$	
ChWy($r2$)Iδ_+(X1,2)	ii	$(\gamma_-)u/2<T<[\alpha_2+(\alpha_2^2+4)^{1/2}]u/2$ γ_- as in Table I of Part II	$T=[\alpha_2+(\alpha_2^2+4)^{1/2}]u/2$	$T=(\gamma_-)u/2$	

will be examined in this section for the case $L=\text{const}$ (W-type solutions and the M-type solutions mentioned above), leaving the cases with $L=L(T)$ for Secs. VII and IX. Throughout this section the coordinate choice II(33) will be used and, unless stated otherwise, it will be assumed that the equation of state has been chosen [choice of $\Theta(T)$] so that the fluid does not bounce in such a way as to avoid $\Pi=0$.

At $\Pi(T,r)=0$, which corresponds in general to finite coordinate values T and r, the metric coefficients H and R diverge as Θ tends to a constant value, which will be different for each class of comoving observers because $\Pi(T,r)=0$ does not coincide with a surface Σ_T. Though in the case of the M-type solutions mentioned above, where $\Pi=0$ becomes $T=0$, Θ has the same "terminal" value $\Theta(0)$ for all comoving observers. Hence, in all these cases, space-time is inextendible beyond this boundary, and from Eq. II(7), H behaves asymptotically along an arbitrary surface Σ_r labeled by $r=r_1$, as

$$H\rightarrow\exp[\tfrac{1}{3}\Theta(T_1)],\tag{7}$$

where T_1 is the time coordinate value satisfying $\Pi(T_1,r_1)=0$. This form of H can be described as a sort of "asymptotically de Sitter" behavior. As $H\rightarrow\infty$ in (7), $\tau\rightarrow\infty$, and so comoving observers reach $\Pi=0$ in their infinite future (or past, depending on the sign of Θ). Although Θ and τ are in general different for $r\neq r_1$, the same qualitative behavior of H, as shown in Eq. (7), occurs for all comoving observers reaching $\Pi=0$.

This asymptotically de Sitter behavior can be connected to the fact that the term $L(fh)^2$ in the right-hand side of Eq. II(16a) plays the role of a sort of position-dependent "cosmological constant." And so, when $L>0$, all these terms, except $L(fh)^2$, vanish asymptotically (as $\tau\rightarrow\infty$ and $H\rightarrow\infty$ for finite r). However, if $L<0$, Eq. II(16a) implies that H and R are necessarily bounded for finite r. Also, for M-type solutions having $L=0$ [except NMcV($r2$)(X1,2) and its charged version], H, and R diverge for finite r only if $|T|\rightarrow\infty$, thus if Θ is chosen (i.e., choice of equation of state) so that $\Theta\rightarrow\text{const}$ as $|T|\rightarrow\infty$, the asymptotically de Sitter behavior associated with Eq. (7) occurs (see Appendix A for an example).

The state variables ρ, p, and q (and thus, curvature scalars such as $\mathscr{R}$ or $\mathscr{R}_{\alpha\beta}\mathscr{R}^{\alpha\beta}$) remain bounded as $\Pi\rightarrow0$. From the field equations II(14b), II(16), II(20), and II(21), for solutions with $L>0$, the state variables approach the following asymptotic values:

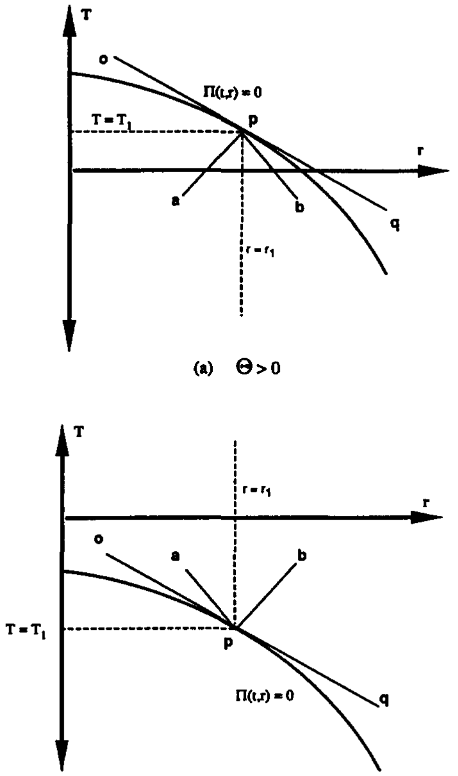
$$4\pi q\rightarrow0,\tag{8a}$$

$$8\pi\rho\rightarrow\Theta^2/3-3(fh)^2L,\tag{8b}$$

$$8\pi p\rightarrow-\Theta^2/3+(fh)^2L,\tag{8c}$$

as $\Pi\rightarrow0$. These values are finite in general, indicating that $\Pi=0$ is not a singular boundary, but rather a sort of coordinate compactification bringing points at infinity into finite coordinate values. This "blowing up" of H and R in the infinite past or future of the comoving observers, together with the asymptotically de Sitter form of the Hubble scale factor in Eq. (7), justifies regarding the evolution of the fluid near $\Pi=0$ as a sort of "inflationary-like phase." This situation occurs under perfectly regular, though unphysical, conditions. In fact, by looking at Eqs. (8b) and (8c), as $\Pi\rightarrow0$, the weak energy condition is violated, since $p+\rho\rightarrow-L(fh)^2/4\pi<0$. Equations (8) with $L=0$ also apply to the M-type solutions NMcV($r2$)(X1,2) and their charged versions, and in general to all M-type solutions in which $\Theta\rightarrow\text{const}$ as $|T|\rightarrow\infty$. In the cases with $L=0$, one has the more acceptable asymptotic limit $p+\rho\rightarrow0$.

Since τ diverges as $\Pi\rightarrow0$, this boundary corresponds to infinite values of the affine parameter of timelike geodesics reaching it. Regarding the completeness of null geodesics, as $H\rightarrow\infty$ for finite r, one has $dv\approx H[(L^{1/2}/h_1)dt+dr]$, which is not an exact differential and thus, the integral in II(55) cannot be computed approximately as it was done in Part II for the boundaries $H=0$ and $Q=0$. However, since R also



(b) Θ < 0

FIG. 8. Light cones near the boundary $\Pi = 0$. Let $\Pi(T_1,r_1) = 0$, so that $\Pi = 0$ is in the infinite future (a) or infinite past (b) of a comoving observer labeled by $r = r_1$. The light cones of this observer near $\Pi = 0$ are represented by the line segments ap and bp, whose slope in the (T,r) coordinates is given by $[dT/dr]_{\rm null}$ [Eq. (9)]. From the construction of this figure, it follows that the boundary $\Pi = 0$ is spacelike at (T_1,r_1) if the absolute value of $[dT/dr]_{\rm null}$ is larger than the slope of the line segment opq [i.e., condition (10)].

diverges as $\Pi \to 0$, and as a result of the invariant characterization of R in spherically symmetric solutions, one would expect these curves to be complete if R diverges along them.[31] Information about the conformal structure of this regularity boundary follows from the study of null geodesics near coordinate values satisfying $\Pi = 0$. As $H \to \infty$ for finite r, Eq. II(56) with the coordinate choice II(33) becomes approximately

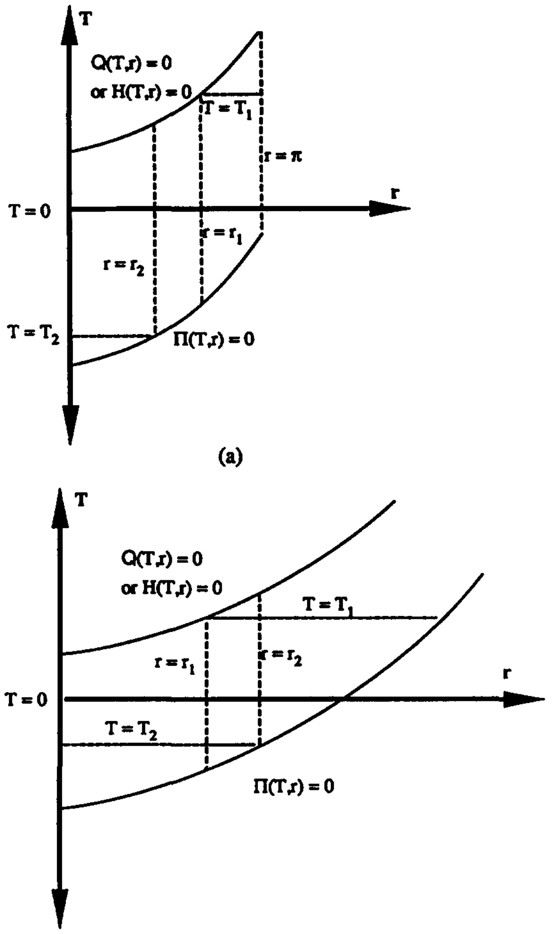
$$\left[\frac{dT}{dr}\right]_{\rm null} \approx \pm \frac{\Theta(T_1)}{3} \frac{h(r_1)}{L^{1/2}}, \qquad (9)$$

where $T = T_1$ and $r = r_1$ satisfy $\Pi(T_1,r_1) = 0$. Since Eq. (9) provides the slopes of the light cones along $\Pi = 0$ in a (T,r) coordinate diagram, the conformal structure of this boundary follows by comparing these slopes with the slope of the boundary $\Pi = 0$ in these coordinates, bearing in mind, of course, that the "interior" of the light cones is given by the "vertical" direction along the timelike world lines of comoving observers ($r = $ const). This is illustrated in Fig. 8. The boundary $\Pi(T,r) = 0$ becomes in (T,r) coordinates a constraint of the form $T = -X + a_0$, where $X = \int h^2\, dy = \int f h^2\, dr$ is given by Eqs. I(25) and a_0 is a constant whose form depends on the parameters of the solutions (see Table II). Hence if $\Theta < 0$ so that $\Pi(T_1,r_1) = 0$ is in the past of the comoving observer $r = r_1$ [see Fig. 8(a)], this boundary will be spacelike if the condition

$$|\Theta(T_1)|/3L^{1/2} \geqslant f(r_1)h(r_1) \qquad (10)$$

holds (and null with the equality sign). It is not possible to say whether this condition will be satisfied in general, independently of the choice of equation of state (choice of Θ) and parameters (a,b,c) and $k = 0, \pm 1$, or all along the boundary. However, in the case of the Wyman solution, which presents this boundary, condition (10) holds for all comoving observers if $|\Theta(T)| > 0$ for all T (i.e., the fluid does not bounce), and so in this case $\Pi = 0$ is everywhere spacelike. This fact will be shown in Appendix A. For CF solutions, $\Pi = 0$ is spacelike (see Sec. IX), and for the solutions NMcV$(r2)(X1,2)$ and their charged versions, this boundary (whose coordinate representation is $T = 0$) is spacelike or null (see Appendix B).

The effect of the occurrence of the boundary $\Pi = 0$ in the global properties of the solutions can be discussed in terms of the topology of the surfaces Σ_T and the topological embedding (6). As $\Pi = 0$ corresponds to infinite values of R for r finite, this boundary cannot occur in solutions of type (i) and (iii). Hence, in a given type (ii) or (iv) solution $\mathscr{M}$, the surfaces Σ_T might (a) extend throughout the full regular range of r, (b) reach only one of the boundaries $\Pi = 0$, $H = 0$, or $Q = 0$, or (c) extend between $\Pi = 0$ and either



(b)

FIG. 9. Domain of regularity of solutions presenting the boundary $\Pi = 0$. As in Figs. 1 and 3, surfaces Σ_T and Σ_r are represented by horizontal and vertical dotted lines. The thick curve below the figures could represent either the boundary $Q = 0$ or $H = 0$, depending on whether the solutions are of type (ii) or (iv).

one of $Q=0$ or $H=0$. Assuming that conditions II(45) hold at $r=0$, the cases $k=1$ and $k=0,-1$ will be considered separately.

A. Case $k=1$

If Σ_T does not reach either boundary [$T=0$ in Fig. 9(a)], it is homeomorphic to S^3 and complete [with an embedding diagram as in Fig. 2(a)]. If Σ_T extends from either one of the centers ($r=0$ or $r=\pi$) to the boundary $\Pi=0$ [$T=T_2$ in Fig. 9(a)], it has topology $\mathbb{R}^3$, since $H\rightarrow\infty$ as $\Pi\rightarrow 0$ and from Eq. (3), ζ diverges if the upper limit of integration in (3) is a value of r satisfying $\Pi(T,r)=0$ [$r=r_2$ in Fig. 9(a)]. As $\mathscr{M}$ is regular at $\Pi=0$, Σ_T is in this case geodesically complete having an embedding diagram as in

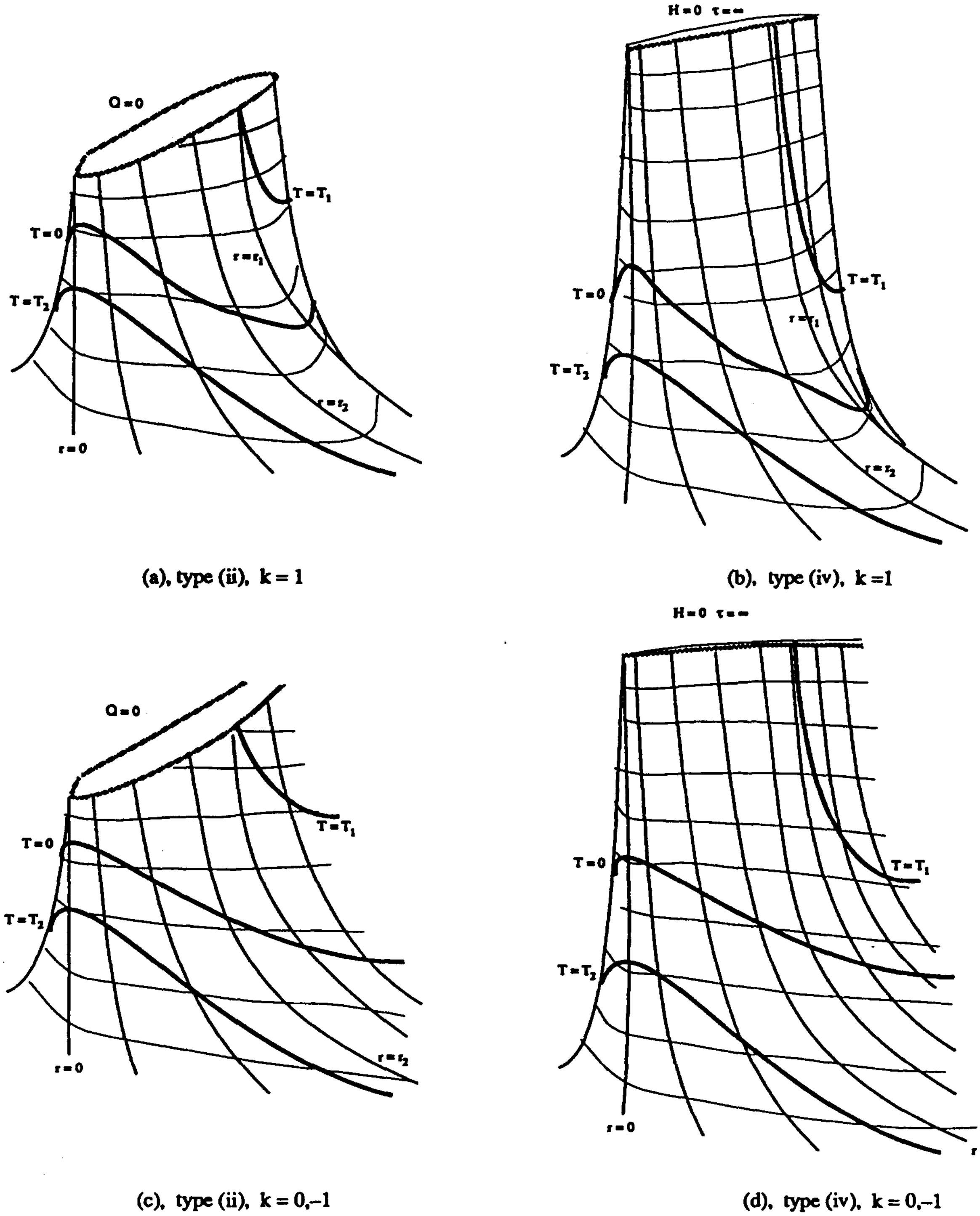


FIG. 10. Representation of the embedding of $\mathscr{M}$ for solutions presenting the boundary $\Pi=0$. As in Figs. 5–7, horizontal and vertical profile curves denote surfaces Σ_r and surfaces of constant τ. In particular, the boundary $\Pi=0$ can be associated with the exponential form of the vertical profile curve $\tau(R)$ (fixed r) in the lower part of the figures. Various surfaces Σ_T of Fig. 9 are depicted as thick, tilted curves. In (a) and (b), $\mathscr{M}$ is homeomorphic to $S^3\times\mathbb{R}$, even though the topology of the complete surfaces Σ_T (for example, $T=T_1$ and $T=T_2$) changes from S^3 to $\mathbb{R}^3$. In (c) and (d), $\mathscr{M}$ is homeomorphic to $\mathbb{R}^4$, as the profile curve $R(r)$ (τ fixed) does not close around a second center.

Fig. 4(c) with $r=\infty$ replaced by $r=r_2$. As in solutions without $\Pi=0$, the topology of surfaces Σ_T within $\mathscr{M}$ changes as surfaces Σ_T reach the boundaries $H=0$ and/or $Q=0$, and now $\Pi=0$.

As shown by the representation of $\mathscr{E}(\mathscr{M})$ of Figs. 10(a) and 10(b), $\mathscr{M}$ looks like a sort of conelike object with a hyperboloidal $\tau(R)$ profile and is clearly homeomorphic to $S^3\times\mathbb{R}$. A given Σ_T reaching $\Pi=0$ [say, at coordinates (r_2,T_2)] stretches towards $\tau\rightarrow\infty$, and so appear as a "tangent plane" of this hyperboloidal shape, avoiding the world lines of comoving observers with $r_1<r<\pi$ [see Figs. 10(a) and 10(b)]. The form of $\mathscr{E}(\mathscr{M})$ near $\Pi=0$ (as $\tau\rightarrow\infty$), as depicted in Figs. 10(a) and 10(b), is qualitatively analogous to the shape of de Sitter space-time as a hyperboloid when embedded in $\mathbb{R}^5$ (see Ref. 8). Such a qualitative resemblance fits the "asymptotically de Sitter" behavior associated with Eq. (7), and as shown by Krasinski,[3] de Sitter space-time can be foliated by spacelike slices whose topology changes (from S^3 to $\mathbb{R}^3$) in a similar way as the topology of the Σ_T reaching $\Pi=0$ changes in Figs. 10(a) and 10(b). This situation will be further commented in Sec. IX.

B. Case $k=0,-1$

Some of these solutions might be homeomorphic to $S^3\times R$, depending on whether another center (besides $r=0$) exists, as in case (5a), and the same arguments discussed for the case $k=1$ apply. For solutions corresponding to the cases given by Eqs. (5c) and (5d), conditions II(45) only hold at $r=0$, though there might not exist surfaces Σ_T which avoid $\Pi=0$ or the singular boundaries [i.e., extending from $r=0$ to $r\rightarrow\infty$, see Fig. 9(b)]. Surfaces Σ_T extending between $r=0$ and $\Pi=0$ [$T=T_2$ in Fig. 9(b)] are geodesically complete and homeomorphic to $\mathbb{R}^3$, their embedding diagram being analogous to that depicted in Fig. 4(c) with $r\rightarrow\infty$ replaced by $r\rightarrow r_2$, where r_2 satisfies $\Pi(T_2,r_2)=0$. On the other hand, surfaces Σ_T reaching either one of the singular boundaries $Q=0$ or $H=0$ [see Fig. 9(b)] are geodesically incomplete, being either homeomorphic to $\mathbb{R}^3$ if they extend to $r=0$, or to $S^2\times\mathbb{R}$ if they reach $\Pi=0$ [$T=T_1$ in Fig. 9(b)].

As shown by the representations of $\mathscr{E}(\mathscr{M})$ of Figs. 10(c) and 10(d), $\mathscr{M}$ is in all cases homeomorphic to $\mathbb{R}^4$. As with the case $k=1$, the representation of $\mathscr{E}(\mathscr{M})$ has a hyperboloidal profile with those surfaces Σ_T reaching $\Pi=0$ being "tangential" to it. However, the $\mathbb{R}^4$ topology occurs because $\mathscr{M}$ does not "close" around a second center, and so just as in solutions with $|\Pi|>0$ (see Figs. 5–7), the singular boundaries $Q=0$ and $H=0$ appear as a "crack" and a "slit" on top of $\mathscr{E}(\mathscr{M})$. Thus, in these cases, the "asymptotically de Sitter" behavior of Eq. (6) does not match a de Sitter-like topology $S^3\times\mathbb{R}$.

C. Particle horizons

As mentioned before, causal curves reaching $\Pi=0$ must be complete. Hence this boundary can be identified as a coordinate representation of a (future/past) regular null infinity surface $\mathscr{I}_\pm$ ("scri" plus/minus) associated with the regular infinite (future/past) of null geodesics and comoving observers. If this boundary is spacelike, then following the discussion in Sec. 5.2 of Hawking and Ellis,[8] one must have particle horizons similar to those arising in de Sitter space-time. This situation is illustrated in the conformal diagram of Fig. 11, which shows this boundary in the case in which it is spacelike. Table II lists those solutions of Tables III and VI of Part I and Table I of Part II in which $\Pi=0$ occurs. The conformal structure and global view of this boundary is discussed for the case of the Wyman solution in Appendix A.

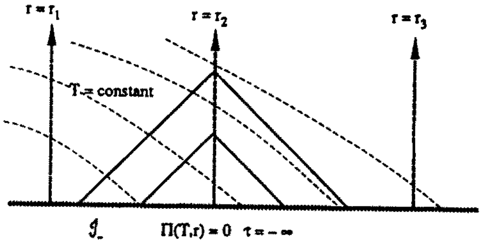


FIG. 11. Particle horizons and the boundary $\Pi=0$. This figure displays a conformal diagram of the boundary $\Pi(T,r)=0$, in the case in which it is a spacelike boundary at the infinite past of causal curves (i.e., a $\mathscr{I}_-$ surface). World lines of comoving observers and surfaces Σ_T are depicted as vertical dotted lines and dotted curves, respectively. The past light cones of the observer $r=r_2$ are shown, extending all the way towards $\Pi=0$ but without reaching $r=r_1$ or $r=r_3$, observers lying outside the particle horizon of $r=r_2$. The boundary $\Pi=0$ need not be always spacelike (see Appendix A).

V. SOLUTIONS WITH "WORMHOLES" AND A SINGULAR NULL $\mathscr{I}$

So far, it has been assumed that "regularity at the center," defined by conditions II(45), holds at $r=0$. In order to comply with this restriction, it has been necessary to exclude those combinations of parameters (a,b,c) in I(23) and I(25) which would make h and X (and thus H) unbounded as $r\rightarrow 0$. Specifically, these parameter combinations, with their corresponding forms of h and X, are the following:

$$a,b\neq 0,\quad c=0,\quad \Delta=b^2,$$

$$\bar{h}_{(1)}=[y(ay+2b)]^{-1/2},\quad \bar{X}_{(1)}=\frac{1}{2b}\ln\frac{ay}{(ay+2b)},\tag{11a}$$

$$a=c=0,\quad b=|\Delta|^{1/2},$$

$$\bar{h}_{(2)}=[2by]^{-1/2},\quad \bar{X}_{(2)}=(1/b)\ln|2by|^{1/2},\tag{11b}$$

$$b=c=\Delta=0,\quad a>0,$$

$$\bar{h}_{(4)}=a^{-1/2}y^{-1},\quad \bar{X}_{(4)}=-(ay)^{-1},\tag{11c}$$

where the bar on top of h and X will distinguish these functions from the cases in which they are bounded as $r\rightarrow 0$. The properties of the locus $r=0$ in the cases presented above, together with its effects on the global structure of the surfaces Σ_T, will be discussed in this section. As in previous sections, the t-parameter T will be chosen as the time coordinate [coordinate choice II(33)].

Although H diverges as $r\rightarrow 0$ in all cases with h and X given by (11), the metric coefficient R might be finite or might diverge as $r\rightarrow 0$, depending on the value of $k=0,\pm 1$ and the specific form of H. However, in all these cases, conditions II(45) do not hold at $r=0$ and thus this locus does not mark the world line of a regular center [fixed point of the orbits of SO(3)], but the time evolution of a two-sphere [of infinite proper radius if $R(0,T)$ diverges]. If a given surface Σ_T extends as far as $r=0$, ζ evaluated from Eq. (3) diverges as $r\to 0$, and thus points at $r=0$ are located at an infinite affine parameter distance along Σ_T. Hence the locus $r=0$ marks a boundary beyond which the space-time manifold $\mathscr{M}$ cannot be extended.

The next question concerns the behavior of curvature scalars as $r\to 0$. In particular, the behavior of $\mathrm{Q}(T,r)$ as $r\to 0$ follows from Eq. I(24a), which can be rewritten as

$$T+\overline{X}=\int\frac{dW}{[Q(W)]^{1/2}}=\int\frac{dW}{W[\mathrm{Q}(W)]^{1/2}},\qquad(12)$$

where Q and Q are given by Eqs. I(21b) and II(25a), respectively, and X is any of the forms of Eqs. (11). As $r\to 0$, X in the left-hand side of (12) diverges, and thus the denominator Q in the right-hand side of this equation must vanish. Since $W=h/H$ is, in general, finite as $r\to 0$, there are three possibilities concerning the vanishing of the product $W[\mathrm{Q}(W)]^{1/2}$:

$$W(T,0)>0\ \Rightarrow\ \mathrm{Q}(T,0)=0,\qquad(13a)$$

$$W(T,0)=0\ \text{and}\ \mathrm{Q}(T,0)=0,\qquad(13b)$$

$$\mathrm{Q}(T,0)>0\ \Rightarrow\ W(T,0)=0.\qquad(13c)$$

The case (13a) comprises all solutions in which $L\neq 0$ irrespective of the form of h and X in (11), since $h/H=\Pi/X$, and the functions Π and Ξ appearing in this quotient are of the same order in r (see Tables III and VI of Part I). The case (13b) corresponds to h and X as in (11c) and $L=0$. The case (13c) consists only of the solutions NMcV($r2$)(X1,2) and their charged versions (see Appendix B). Other M-type solutions with h and X as in (11a) and (11b) belong to case (13a) (see Table I).

In the cases (13a) and (13b), the fact that $\mathrm{Q}(0,T)=0$ implies that the terms $d\Theta/d\tau=(\Theta/3)\mathrm{Q}^{-1/2}\,\partial\Theta/\partial T$ and $\mathscr{A}$ in the Raychaudhuri equation II(20) and in II(28) must diverge (and thus $p\to\infty$), just as in the case of the FD singularity $\mathrm{Q}(r,T)=0$ [Eq. II(49)]. Using Eqs. II(14a) and II(26), it can be verified that matter-energy and charge densities, ρ and q, tend to finite values as $r\to 0$ for all forms h and X given by Eqs. (11), except (11c) with $L\neq 0$, in which case both of these quantities diverge as well. Therefore, in most solutions belonging to cases (13a) and (13b), the boundary $r=0$ is singular, and one has $|dp/d\rho|\to\infty$ as $r\to 0$ [just as near the FD singularity $\mathrm{Q}(t,r)=0$], and thus the strong and dominant energy conditions are violated at this asymptotic limit.

On the other hand, for solutions corresponding to the case (13c), all state variables, and thus curvature scalars, are bounded as $r\to 0$, and thus the boundary $r=0$ is regular. In fact, for this case $\mathscr{A}$ vanishes, while q, ρ, and p take the limiting values

$$q\to 0,\qquad(14a)$$

$$8\pi\rho\to\Theta^2/3,\qquad(14b)$$

$$8\pi p\to-\Theta^2/3,\qquad(14c)$$

which, regardless of the choice of Θ (choice of equation of state), leads to the asymptotical limit $p+\rho\to 0$ (see Appendix B).

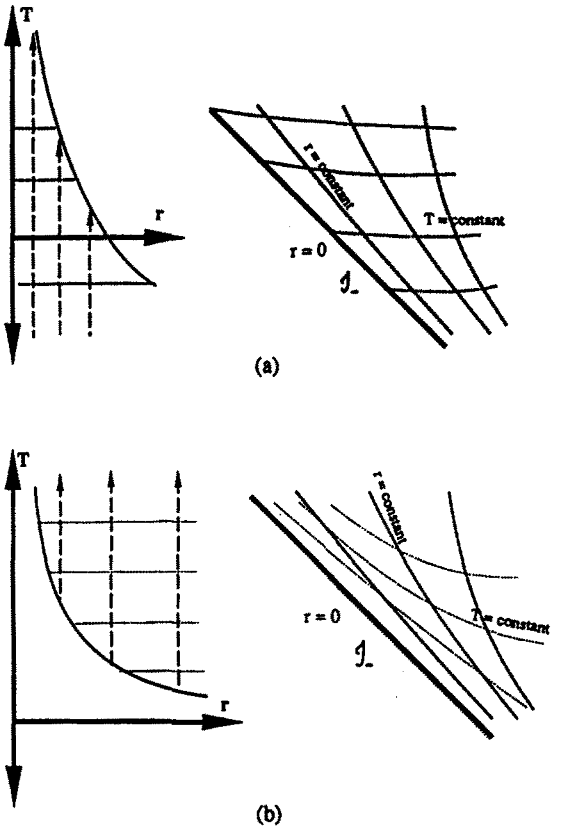


FIG. 12. The null boundary at $r=0$. A conformal diagram of the null boundary at $r=0$, arising in the solutions of Sec. V, is displayed together with surfaces Σ_T and Σ_r. In (a), $r=0$ is the null limit of timelike surfaces Σ_r with $r\to 0$. In this case, all spacelike surfaces Σ_T reach $r=0$. In (b), $r=0$ is the null limit of timelike surfaces $\Sigma_r(r\to 0)$ and spacelike surfaces $\Sigma_T(T\to\infty)$. Both situations described in (a) and (b) can occur within a single solution (see Appendix B).

From Eq. II(56), using the coordinate choice II(33), one has $[dT/dr]_{\text{null}}\to\pm\infty$ as $r\to 0$, and since $r=0$ is a "vertical" line in a (T,r) coordinate diagram, this boundary is necessarily a null hypersurface arising as a null limit of timelike hypersurfaces Σ_r generated by the world lines of comoving observers labeled by values of r arbitrarily close to $r=0$ [see Fig. 12(a)]. Since the function Ω in II(58) diverges as $r\to 0$, the integral in II(55) diverges and so does the affine parameter ϑ along null geodesics reaching this boundary. Therefore, $r=0$ is a null "future (or past) null infinity surface" $\mathscr{I}_\pm$, which in the cases (13a) and (13b) above is singular, while in the case (13c) is regular.

The existence of the null $\mathscr{I}$ at $r=0$ (whether regular or singular) depends on the parameters a, b, and c in Eq. I(25), and thus does not exclude the existence of other boundaries, such as $H=0$, $\mathrm{Q}=0$, and/or $\Pi=0$. However, the fact that H and ζ diverge as $r\to 0$ means that this null boundary affects the topology of surfaces Σ_T in solutions having h and X as in (11). This situation will be discussed below for the cases $k=1$ and $k=0,-1$.

If $k=1$, $R(T,\pi)=0$ and conditions II(45) hold at $r=\pi$. Hence, these solutions have a regular center [fixed point of SO(3)] at $r=\pi$, and so surfaces Σ_T embedded in $\mathbb{R}^4$ will look like those corresponding to the cases $k=0,-1$ of Eqs. (5c) and (5d) (Fig. 4). However, instead of opening from "left to right" [from $r = 0$ to $r \to \infty$, see Fig. 4(c)], in this case Σ_T opens from "right to left" (from $r = \pi$ to $r \to 0$). Depending on whether the Σ_T are complete or not (i.e., if they reach $r = 0$ or the singular boundaries $H = 0$ and/or $Q = 0$), they will be homeomorphic to $\mathbb{R}^3$ or $S^2 \times \mathbb{R}$. Following the arguments of Sec. III, the space-time manifold in these cases is homeomorphic to $\mathbb{R}^4$, having an embedding diagram under (6) as in Fig. 7 with $r = 0$ and $r = \infty$ relabeled as $r = \pi$ and $r = 0$, respectively.

If $k = 0, -1$ with $\zeta \to \infty$ as $r \to \infty$, as in cases (5c) and (5d), then there is no regular center along any of the surfaces Σ_T, all of them being now homeomorphic to $S^2 \times \mathbb{R}$. This is so, even if there are complete surfaces Σ_T (i.e., they extend along $0 < r < \infty$ without reaching any singular boundary). Thus in this case, all surfaces Σ_T embedded in $\mathbb{R}^4$ have the wormhole shape depicted in Fig. 13, analogous to the "Einstein–Rosen bridge" in the Schwarzschild and Reissner–Nordstrøm solutions (see Sec. 31.6 of Ref. 32). Hence the space-time manifold in these cases is homeomorphic to $S^2 \times \mathbb{R}^2$, as illustrated in the representation of $\mathscr{C}(\mathscr{M})$ of Fig. 14. The existence of particular cases of the Tolman–Bondi dust solution with this type of topology has been recently reported by Hellaby.[33]

As mentioned in the previous section, the boundaries $H = 0$, $Q = 0$, or $\Pi = 0$ are usually characterized in (T,r) coordinates as constraints of the form $T = -X + a_0$, where X is given by I(25) and a_0 is a constant whose form depends on the parameters of the solution (see Table I of Part II). Thus, if X takes one of the forms in (11), the domain of regularity of the solution might be such that the surfaces Σ_T

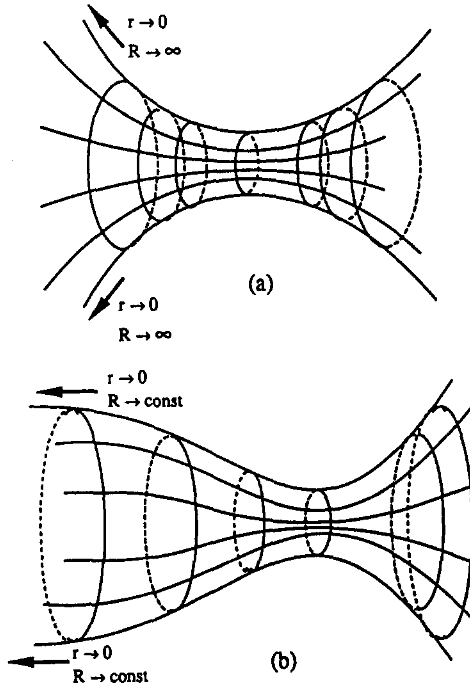


FIG. 13. Representation of the embedding of surfaces Σ_T with wormhole shape. Surfaces Σ_T of solutions homeomorphic to $S^2 \times \mathbb{R}^2$, discussed in Sec. V, do not have a regular center. Hence, if these three-surfaces extend towards $r = 0$, the 2-D representation of the embedding given by Eq. (4) will look as in these figures, depending on whether R diverges (a) or tends to a finite value (b) as $r \to 0$. Though these solutions might not have complete Σ_T surfaces, the latter would necessarily be homeomorphic to $S^2 \times \mathbb{R}$.

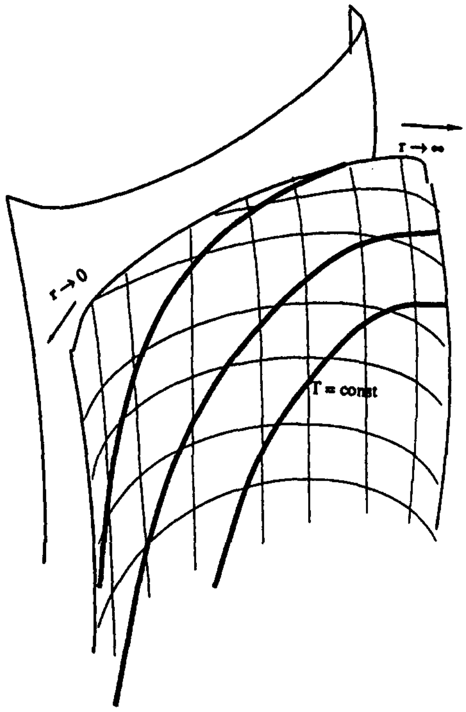


FIG. 14. Representation of the embedding of $\mathscr{M}$ for solutions homeomorphic to $S^2 \times \mathbb{R}^2$. This representation of $\mathscr{C}(\mathscr{M})$ is equivalent to those shown in Figs. (5)–(7) and (10). Now for the case of solutions lacking regular centers, see Sec. V. As in previous representations, horizontal and vertical profile curves denote surfaces of constant τ and r, respectively [see Fig. 5(a)]. These surfaces are connected, and their appearance as disconnected curves is only an effect of the representation. Surfaces Σ_T are depicted as thick, tilted curves. The $S^2 \times \mathbb{R}^2$ topology of $\mathscr{M}$ follows from the fact that horizontal profile curves $R(r)$ (τ fixed) fail to "close" due to the lack of centers in $\mathscr{M}$. On top of the figure, depending on whether the solution is of type (ii) or (iv), one could have a FD singularity or an AD big bang.

only reach $r = 0$ as $T \to \infty$. In this case, which is illustrated in Fig. 12(b), $r = 0$ is the null boundary arising as a null limit of timelike hypersurfaces (surfaces Σ_r generated by the world lines of comoving observers with r close to zero) and spacelike hypersurfaces (surfaces Σ_T with $T \to \infty$).

As far as I am aware, the fact that $r = 0$ marks a null $\mathscr{I}$ (which in most cases is singular) in ChKQ solutions having h and X as in (11) has never been noticed or reported in the literature. In particular, McVittie's solution[9] (see Appendix B) presents this feature, but has been overlooked by authors studying it (see Appendix B for references). Mashhoon and Partovi[12] noticed that p diverges as $r \to 0$ in solutions ChMcV($r2$,$r2$)(X 1,2) having h and X as in (11a) and (11b) and for the uniform-density solution ChMcV(UD)($r2$) (see their Sec. VI and their Appendix C). However, they mistook the singular null $\mathscr{I}$ arising in the former solutions (see Appendix C in Part II) with the FD singularity $Q(r,T) = 0$ (which, incidentally, also arises). The null (and possibly singular) $\mathscr{I}$ at $r = 0$ and the fact that all surfaces Σ_T are homeomorphic to $S^2 \times \mathbb{R}$ (i.e., wormholes) is also significant in the study of gravitational collapse of fluid spheres modeled by solutions with h and X as given by Eqs. (11). This will be discussed in the following section.

VI. COLLAPSING FLUID SPHERES WITHOUT A CENTER

The forms of h and X given by Eqs. (11) do not contradict with matching the solutions discussed in the previous section to Schwarzschild or Reissner–Nordstrøm solutions. In fact, such a matching has been done previously for the McVittie solution[9] [NMcV($r2$)($X2$) with $c = 0$], in which h and X are as in (11b), by McVittie[10] and Knutsen.[11] However, these authors failed to notice that the locus $r = 0$ is no longer the world line of a center but a null $\mathscr{I}$ surface, and so they did not study the effects of this null boundary (which is regular in this case) on the global structure of the collapsing configuration. These effects do not occur in the usual spherically symmetric collapse in which $r = 0$ marks the "center" of the collapsing sphere.

If h and X are given by Eqs. (11), the standard collapse picture with $r = 0$ marking a regular center, as described in Sec. XI of Part II, can only happen if the interior fluid region is matched to Schwarzschild or Reissner–Nordstrøm at a surface Σ_r in such a way that $r > 0$ throughout the fluid region (i.e., h and X are bounded for all r). This can only be done for the cases having a regular center (either at $r = \pi$ or at $r = \infty$), and by relabeling the coordinate r, these "left-side" matchings coincide with the ones discussed in Sec. XI of Part II.

If the fluid region contains the locus $r = 0$, it has no regular center and this fluid region plus the vacuum Schwarzschild or Reissner–Nordstrøm region, matched along a surface Σ_r labeled by $r = r_0$, combine into a hybrid space-time manifold homeomorphic to $S^2 \times \mathbb{R}^2$ having spacelike slices with a wormhole $S^2 \times \mathbb{R}$ topology as in a Kruskal diagram with "two sheets" (i.e., without fluid region, see Refs. 33–36). Although the intuitive idea of interior cannot be applied to the fluid region, as the latter is observed from the vacuum region (say, from a static frame), it appears as a two-sphere (i.e., the surface of the fluid sphere) of finite proper radius $R_0 = R(T, r_0)$ whose proper time evolution is governed by Eq. II(73a). Hence, if one is only interested in studying the kinematical evolution of this two-sphere in the Schwarzschild or Reissner–Nordstrøm geometries (whether it bounces or collapses), then the global structure of the fluid region can be ignored. Otherwise, the effects due to the lack of a center in the fluid region must be brought into consideration.

If the fluid region is a type (ii) solution then, as shown by Figs. 15(a) and 15(b), the world lines of comoving observers and null geodesics terminate at the FD singularity, while the surface becomes singular at $R_0 = f_0 h_0 / A$, where A is a root of Q given by I(21b). If the solution is of type (iv), then as shown by Figs. 15(c) and 15(d), comoving observers along the surface $r = r_0$ collapse into the null L singularity, while the remaining comoving observers $(0 < r < r_0)$ evolve towards the AD big bang in their infinite future. In the latter case, one has a similar situation as described in Figs. 9 and 10 of part II, in which the proper volume of the "surface" of the sphere $(4/3\pi R_0^3)$ vanishes, while proper local volumes along interior layers $(\sim H^3)$ remain nonzero. Though, if the fluid region has no regular center, the proper volume of the orbits of SO(3) corresponding to these internal layers never vanishes (i.e., there is no center) and could even be infinite (if $R \to \infty$ as $r \to 0$). This situation reveals to what degree the labels "interior" and "exterior," as applied to the fluid and vacuum regions, lose their intuitive meaning

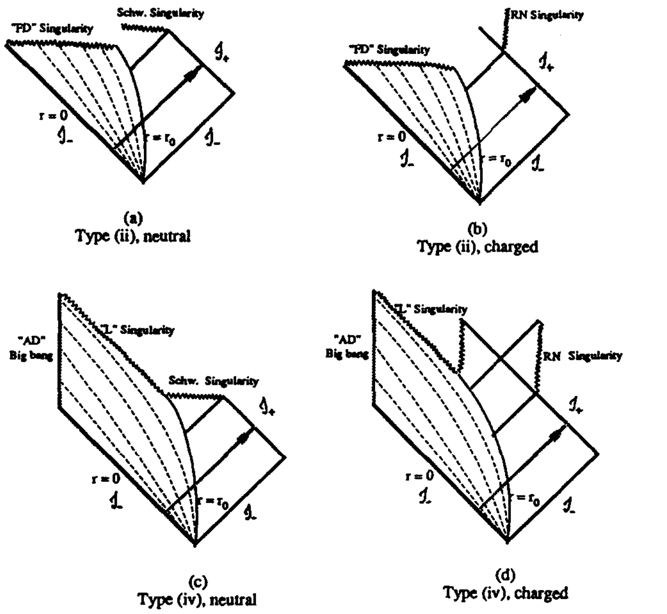


FIG. 15. Qualitative Penrose diagrams of collapsing spheres without a center. These Penrose diagrams are analogous to those introduced in Sec. X of part II, except that now $r = 0$ is not the world line of the center of the sphere (see Sec. VI and Sec. VIII of part II), but a null boundary at infinity (a null $\mathscr{I}$ surface). In (a) and (b), the world lines of comoving observers (dotted curves) terminate at the FD singularity (compare with Fig. 12 of Part II). In (c) and (d), internal fluid layers, labeled by $0 < r < r_0$, evolve towards the AD big bang in their infinite future. The surface of the sphere (the two-sphere labeled by $r = r_0$, whose time history is denoted by a solid curve) collapses into the null L singularity, as in Figs. 11 and 12 of Part II. The gray arrows in (c) and (d) represent light rays from $r = 0$ reaching the Schwarzschild or Riessner–Nordstrøm regions, so that if this boundary is singular, it behaves as a sort of white hole.

in this case. In fact, the global structure of such a fluid region (as a region of space-time) corresponds to a cosmological type of space-time and not a model of a stellar interior. Thus, observers entering a fluid spheres of type (iv) from the vacuum region (see Fig. 16), not only could survive the black hole (the L singularity), but could experience the "science fiction" effect of entering a "star" which is in its own right "another universe."

Although a fluid sphere without a center would appear to observers in the vacuum region as an object of stellar dimensions (R_0 is finite and it could even be arbitrarily small), the fact that the fluid region inside this object is not compact could be detected locally by these observers. The latter would be extremely puzzled as they detect red or blue shifts of cosmological proportions in photons emitted by such a star (this situation is illustrated in Fig. 16). If the null $\mathscr{I}$ at $r = 0$ is singular, light rays from it can reach distant observers in the Schwarzschild or Reissner–Nordstrøm region (see Figs. 15 and 16), and so this singular boundary behaves in a similar manner as the white hole in a "two-sheet" Kruskal diagram without a collapsing body.[34–36] Though, the important difference is that this white-hole effect is produced by the fluid region itself, without having to justify the absence of matter or to impose any peculiar topological identification.[33,36]

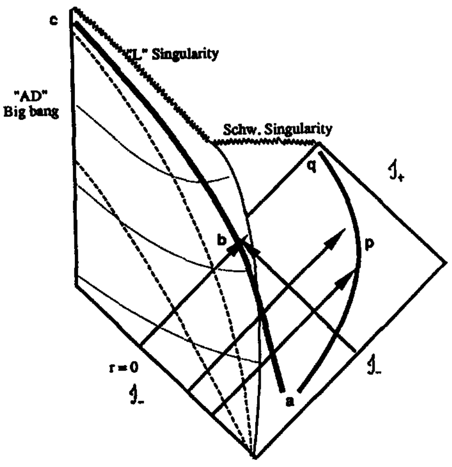


FIG. 16. An observer falling into a sphere without a center. The history of such an observer is depicted by the timelike (nongeodesic) curve abc. At point "a" the observer is in a static frame (world line apq) in the Schwarzschild region, say, in a planet in circular orbit ($R = \text{const} > 2m$) around the sphere. The latter is seen in a as a spherical star with finite radius. As the journey progresses, the observer goes inside of this star, and at point b detects radiation coming from the null infinity surfaces $\mathscr{I}_+$ of the fluid region ($r = 0$) and of the vacuum region. The existence of the former radiation, which could be detected also in a, makes the observer suspect that the fluid region is not the interior of a star but "another universe" (a region with spacial extension of cosmological proportions). As the journey continues, past of b, it becomes impossible for the observer to return to the static frame apq. As the surface of the sphere crosses the horizon $R = 2m$, and collapses into the L singularity, the observer could (as in Fig. 11 of Part II) survive the collapse and dwell inside of the sphere forever (his/her world line is complete, provided he/she avoids the L singularity and heads towards the AD big bang).

VII. SOLUTIONS WITH $L = L(t)$

If L is a second t-parameter, Eq. II(16) and all derivatives with respect to r (or y), such as II(17) or II(24), remain unchanged, with the extra feature that $L = L(t)$. Regarding time derivatives like II(30) in solutions with $\Psi_{(2)} \neq 0$ (the case $\Psi_{(2)} = 0$ is discussed in Sec. IX), the modulus η of the elliptic integral in I(24a) is now a time-dependent function, and so the derivative $\dot{H}/H$ has the following form:

$$\dot{H}/H = \dot{R}/R = \mathrm{Q}^{1/2}[\mathscr{N}\dot{L} - \dot{T}], \tag{15}$$

where

$$\mathscr{N}(t,r) \equiv \left[\frac{\partial \mathrm{F}}{\partial \Psi}\frac{\partial \Psi}{\partial \eta} + \frac{\partial \mathrm{F}}{\partial \eta}\right]\frac{d\eta}{dL}$$
$$= \left[\frac{\xi}{\eta\bar{\eta}^2}\mathrm{E}[\Psi,\eta] + \frac{\eta\xi\sin\Psi\cos\Psi}{\bar{\eta}^2(1-\eta^2\sin^2\Psi)^{1/2}} + \frac{N+X}{\eta}\left[1 - \frac{d(\ln\xi)}{d\eta}\right]\right], \tag{16a}$$

with $\bar{\eta}^2 = 1 - \eta^2$, and

$$\mathrm{E}[\Psi,\eta] \equiv \int [1 - \eta^2\sin^2\Psi]^{1/2} d\Psi \tag{16b}$$

is an elliptic integral of the second kind.[37] However, $\mathscr{N}(t,r)$ must satisfy the integrability condition II(27) leading to the constraint

$$\mathscr{N}' = \mathrm{Q}^{-1/2}, \tag{17}$$

which allows one to determine the form of the elliptic integral (16b). The function $\mathscr{N}(t,r)$ is quite a cumbersome expression, and thus, in order to examine the properties of solutions with $L = L(t)$ and $\Psi_{(2)} \neq 0$, it might be necessary to use Taylor series approximations or numerical methods. Therefore, this section will only aim to point out in which aspects the properties of these solutions might be qualitatively different from (or similar to) those of the M- and W-type solutions discussed so far. A proper investigation of solutions with $L = L(t)$ and $\Psi_{(2)} \neq 0$ is a topic of further research.

The first consequence of having Eq. II(30) replaced with (15) is that the proper time derivative operator II(8) with $U = \mathrm{Q}^{-1/2}$ and the magnitude of the four-acceleration in II(28) [and so the "acceleration" term II(21a) in the Raychaudhuri equation] become modified as

$$\frac{d}{d\tau} = \frac{\Theta/3}{\mathrm{Q}^{1/2}[\mathscr{N}\dot{L} - \dot{T}]}\frac{\partial}{\partial t}, \tag{18a}$$

$$\mathscr{A} = \frac{fh^2}{\mathrm{Q}^{1/2}}\left[\epsilon^2 W^2 - \mu W + LW^{-2} + \frac{\dot{L}W^{-2}}{\mathscr{N}\dot{L} - \dot{T}}\right]. \tag{18b}$$

Equation (18a) indicates that condition II(29) must be in turn modified by demanding now that Θ and the derivatives of the t-parameters must be of the same order if expanded around a value $t = t_0$ such that $\Theta(t_0) = 0$. That is, Θ, T, and L must satisfy

$$\Theta/3 = 0 \Leftrightarrow \dot{L} = 0 \Leftrightarrow \dot{T} = 0, \tag{19}$$

which is a strong restriction on the t-parameters. If either one of the t-parameters is rewritten as H_0 (or R_0) as in Eq. I(29), equations similar to (15)–(17) arise, with $\dot{H}_0/H_0$ replacing either one of $\dot{T}$ or $\dot{L}$. Another aspect of solutions with $L = L(t)$ is that the existence of two t-parameters allows one to impose further conditions on the state variables. These conditions can be set in the form of two localized equations of state (see Sec. VII of Part II), each one on a different surface Σ_r, or in further restrictions on a single surface Σ_r.

The regularity boundary $H = 0$ [Eq. II(48)] is also present in solutions type (i) and (iv) with $L = L(t)$, and is a timelike boundary since $(dt/dr)_{\text{null}}$ also vanishes as $H \to 0$. Since R vanishes for $r > 0$ as $H \to 0$, from the discussion of Sec. X of Part II, the boundary $H = 0$ has also the characterization of an AD big bang if Θ remains finite along it. However, the fact that now $L = L(T)$ [using the coordinate choice II(33)] could prevent the development of an AD big bang and instead result in a standard big bang characterized by $H = 0$ coinciding with a Σ_T along which Θ diverges. This situation can be illustrated by the neutral particular case which follows from Eqs. I(17), I(21b), and I(23) by setting $k = 0$, $\epsilon = \Delta = 0$, $\mu < 0$, $L(T) > 0$ for all T, $a = b = 0$, and $c = 1$. These parameter restrictions lead to H having the form

$$H = l(T)(1 - \mathrm{cn}[T + r^2/2])/\mathrm{cn}[T + r^2/2], \tag{20}$$

where $l(T) \equiv [L(T)/2\mu]^{-1/3}$. This form of H is similar to I(37c), that is, the form of a type (iv) Wyman solution (see Table II of Part II), except that $L(T) > 0$ and $\Theta(T)$ are unspecified functions (t-parameters). If H is given by (20), the boundary $H = 0$ becomes

$$H = 0 \Rightarrow \begin{cases} l(T_0) = 0, & 0 \leqslant r \leqslant r_0, \\ \mathrm{cn}[T + r^2/2] = 1, & r \geqslant r_0, \end{cases} \tag{21}$$

where r_0 satisfies $H(T_0, r_0) = 0$ (see Fig. 17). If this solution is restricted to $0 \leqslant r \leqslant r_0$, matching to a Schwarzschild exterior at $r = r_0$, so that $\Theta(T_0) \to -\infty$ is given by II(74a), then the AD big bang does not develop and one has instead a fluid sphere collapsing into a spacelike standard big-bang singularity. In this configuration, the t-parameter $l(T)$ could be determined by imposing a "localized" equation of state along a surface Σ_r (for example, at the center $r = 0$, see Sec. VII of Part II). If the fluid is unbounded, then as shown by Fig. 17, one has a combination of standard big bang ($T = T_0$, $0 \leqslant r < r_0$), null L singularity ($T = T_0$, $r = r_0$), and AD big bang ($T < T_0$, $r > r_0$). However, a form of $H = 0$ like that of Eq. (21), which allows one to eliminate the AD big bang, might not be possible for all solutions with time-dependent L. And even if it is possible, as shown in this simple particular case, the complicated form of the metric coefficients [especially g_{tt} formed from Eq. (15)] makes it difficult to study these cases in detail. This is not so with conformally flat solutions, which also have two t-parameters but simple metric coefficients. These solutions are discussed in Sec. IX.

For whatever form of the two t-parameters, as in the case of M- and W-type solutions, the converse of condition II(29) does not hold, but now it leads to a more complicated constraint as a result of the fact that the quantity $Q^{1/2}[\mathscr{N}\dot{L} - \dot{T}]$ in (15) does not necessarily vanish if Q vanishes and vice versa. This follows from the forms of $\Psi(W,\eta)$ presented in Tables I and II of Part I. However, from (18b), $\mathscr{A}$ also diverges as $Q \to 0$, and so besides the regularity boundary II(49), solutions with $L = L(t)$ might have an extra regularity boundary defined as the set of (t,r) values such that $[\mathscr{N}\dot{L} - \dot{T}] = 0$. Also, the form of H given in II(43) has $\Pi = \Pi(L,T,r)$ [see Eqs. I(28), I(30), and I(32)] so that, depending on the specific form of $\Pi = 0$,

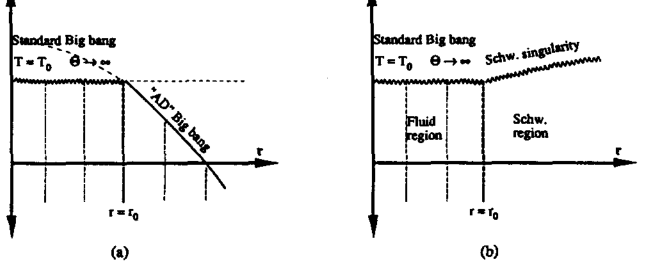


FIG. 17. Standard big bang in solution with $L = L(T)$. (a) displays the boundary $H(T,r) = 0$ for the example of Eqs. (20) and (21). This boundary becomes a standard big bang ($|\Theta| \to \infty$ as $H \to 0$) for $T = T_0$ and $0 \leqslant r < r_0$. For $r \geqslant r_0$, there is a combination of L singularity and AD big bang. In (b), this solution is matched to a Schwarzschild space-time, at $r = r_0$. The L singularity and AD big bang are then replaced by the Schwarzschild singularity. Without such a matching, $H = 0$ must take the form shown in (a). The existence of a standard big bang does not occur in general in solutions with $L = L(T)$.

there might be a regularity boundary of the form II(50) related to the asymptotically de Sitter behavior discussed in Sec. IV. Since a study of the conformal structure of regularity boundaries connected with the vanishing of Q, $\mathscr{N}\dot{L} - \dot{T}$ or Π requires testing the convergence of integrals equivalent to II(53) and II(55), and evaluating equation II(56) near the regularity boundaries, it will not be attempted in this paper.

Finally, the topology of these solutions can also be inferred following the arguments previously applied to M- and W-type solutions. If conditions II(45) hold at $r = 0$, the space-time manifold $\mathscr{M}$ will be homeomorphic to $S^3 \times \mathbf{R}$ if $k = 1$ or if $k = 0, -1$ and R and ζ behave as in (5a). In the cases (5c) and (5d), $\mathscr{M}$ is homeomorphic to $\mathbf{R}^4$. If h and X are given by Eqs. (11), then $\mathscr{M}$ is homeomorphic to $\mathbf{R}^4$ if $k = 1$ or $k = 0, -1$ [case (5a)], or to $S^2 \times \mathbb{R}^2$ if $k = 0, -1$ [cases (5c) and (5d)]. If h and X are given by Eqs. (11), $r = 0$ marks a null boundary, which will necessarily be singular as solutions with $L = L(t)$ obviously belong to the case (13a). From Eqs. II(43), I(28), I(30), and I(32), the functions Ξ and Π are of the same order in r as $r \to \infty$ [or as $r \to 0$ if h and X are given by (11)], and so $\int H\,dr \sim \int h\,dr$ and the convergence of ζ and R as $r \to \infty$ or as $r \to 0$ follows from the forms of h and f (see Table I).

VIII. UNIFORM-DENSITY SOLUTIONS

As mentioned in Part I, uniform-density ChKQ solutions follow as the parameters (a,b,c) appearing in the functions h and $X = \int h^2\,dy$ are restricted as in I(38) and I(39), leading to $\rho = \rho(t)$ given by I(41). Table VII of Part I classifies these solutions, some of which are conformally flat ($\epsilon = \mu = 0$, $\Psi_{(2)} = 0$) and will be examined separately in the next section. Hence this section will be confined to those uniform density solutions with $\Psi_{(2)} \neq 0$, to be denoted henceforth as UD solutions. The coordinate choice II(33) will be used.

As shown in Appendix C using the arguments of Appendix B of Part I, different values of k in Eqs. II(2) do not denote different UD solutions. Hence one can choose the r coordinate so that $f(r) = r$ (case $k = 0$) and, from I(38) and I(39), the functions h and X associated with these solutions are

$$h_{(\mathrm{UD})} = c_0/r, \tag{22a}$$

$$X_{(\mathrm{UD})} = c_0 \ln r, \tag{22b}$$

where $c_0 \equiv b^{-1/2} = \Delta^{-4}$. Since h and X in Eqs. (22) are particular cases of Eq. (11b), UD solutions are really particular cases of the solutions discussed in Sec. V. Hence the locus $r = 0$ is not the world line of a center [fixed point of SO(3)], but a null boundary, a null $\mathscr{I}$ that is singular for UD solutions belonging to cases (13a) and (13b) in Sec. V (see Table I).

Matter-energy density is constant along the surfaces Σ_T, and so Eq. II(16a) reduces to Eq. I(41), expressed as

$$\left[\frac{\Theta H}{3}\right]^2 = \left[\frac{dH}{d\tau}\right]^2 = \frac{1}{H}\left[\frac{8}{3}\pi\rho H^3\right] + c_0^2 L(T) H^2, \tag{23}$$

where $\rho = \rho(T)$. Pressure follows from the Raychaudhuri equation [II(20) and II(21)], which for uniform density solutions takes the form

$$\frac{1}{H}\frac{d^2H}{d\tau^2} = \frac{\Theta^2}{9} + \frac{d}{d\tau}\frac{\Theta}{3}$$
$$= -4\pi(p + (\rho/3)) + c_0^2 L$$
$$+ \frac{c_0^2 \dot{L}}{2Q^{1/2}[\mathscr{N}\dot{L} - 1]}, \qquad (24a)$$

where $\dot{L} = dL/dT$ and

$$c_0^2 Q = 1 - 2\mu c_0^5/R + \epsilon^2 c_0^6/R^2 + L c_0^2 R^2. \qquad (24b)$$

Comparing Eqs. II(18) and (20), the Ricci scalar of the surfaces Σ_T is given by ${}^{(3)}\mathscr{R} = -6c_0 L(T)$, and so it is constant along these hypersurfaces [though a different constant along each Σ_T if $L = L(T)$]. The fact that ${}^{(3)}\mathscr{R}$ is a constant, whose sign is the opposite of the sign of $L(T)$, along the Σ_T does not mean that the latter are hypersurfaces of constant curvature, in the sense of having a Ricci tensor of the form (2a). This fact can be verified by computing the components of Ricci tensor of the Σ_T from the metric (1):

$$\mathscr{R}_{rr} = 2\left[-\frac{\mu c_0^5}{R^3} + \frac{\epsilon^2 c_0^6}{R^4}\right] + {}^{(3)}\mathscr{R} = 2\Psi_{(2)} + {}^{(3)}\mathscr{R}, \qquad (25a)$$

$$\mathscr{R}_{\theta\theta} = \mathscr{R}_{\phi\phi} = \mu c_0^5/R^3 + {}^{(3)}\mathscr{R}, \qquad (25b)$$

where H' and H'' have been eliminated from I(18) and II(24), specialized to UD solutions by I(38) and I(39) (i.e., $W = c_0/R$) and (24b). From Eqs. (25a) and (25b), it follows that a constant Ricci scalar implies a Ricci tensor of the form (2a) only if $\mu = \epsilon = 0$, which implies $\Psi_{(2)} = 0$, that is, such an implication is only valid for conformally flat solutions (see the next section).

The field equations (23) and (24) indicate that the FD singularity associated with the vanishing of $(-g_{tt})^{1/2}$ [$Q = 0$ if $dL/dT = 0$, or the vanishing of (15) if $dL/dT \neq 0$] can also occur in UD solutions of types (i) and (ii), and if it does, it will be a spacelike singularity (from the arguments of Fig. 2 of Part II). Notice that in the case $dL/dT = 0$, in which the last term in (24a) vanishes, p still diverges as $Q \to 0$, because the term $d\Theta/d\tau$ becomes infinite at this boundary. Though in this case, the finite-density singularity $Q = 0$ coincides with a surface of constant R [$R = c_0/A$, where A is a root of Q in I(21b)].

The boundary $H = 0$ [Eq. II(48)] might occur in UD solutions of types (i) and (iv). From Eq. (23), matter-energy density ρ only diverges if $\Theta(T)$, $L = L(T)$, or both diverge, this would happen (if it happens) in a (singular) surface Σ_T. If $T = T_0$ marks this singular Σ_T, then one has along $T = T_0$ a spacelike singularity, which would be the FV singularity (i.e., finite-volume singularity) or of a standard big-bang type, depending on whether $H(r,T_0) = 0$ holds or not (see Sec. X and Fig. 4 of Part II). The occurrence of a standard big bang is only possible for solutions with $L = L(T)$ in a way analogous to the example discussed in Sec. VII [Eqs. (20) and (21)] and depicted by Fig. 17. If L is a constant, then $H = 0$ cannot coincide with $T = T_0$ for more than one surface Σ_r and thus the singularity structure of $H = 0$ (AD big bang, L and FV singularities) is similar to W- or M-type solutions with $\rho' \neq 0$, as discussed in Sec. X and Fig. 5 of Part II. However, unlike the case with $\rho' \neq 0$, ρ and p from (23) and (24) (and thus curvature scalars) are bounded at the AD big bang part of this boundary [i.e., those (T,r) values satisfying $H(T,r) = 0$ in which $\Theta(T)$ and $L(T)$ are bounded]. Although the AD big bang is not a singular boundary, from Eqs. II(56), II(64) to II(66), and Fig. 3 of Part II, $H = 0$ is also in this case a timelike boundary marking the infinite future (or past) of comoving observers and timelike and null geodesics. From (23) and (24), $p + \rho \to 0$ as $H \to 0$ if $dL/dT = 0$.

Surfaces Σ_T in UD solutions, whether they reach the null boundary $r = 0$ or terminate at either one of the boundaries $Q = 0$, $H = 0$, or $\Pi = 0$, are necessarily homeomorphic to $S^2 \times \mathbb{R}$, looking like the embedding diagrams of Fig. 13, and the representation of the embedding of $\mathscr{M}$ given by Eq. (6) looks like Fig. 14, and so the space-time manifold is homeomorphic to $S^2 \times \mathbb{R}^2$. For solutions of types (i) and (iv), the fact that the AD big bang $H = 0$ is a regular timelike boundary does not mean that surfaces Σ_T reaching this boundary "close" as surfaces Σ_T of Figs. 4(a) and 4(c) close around the center $r = 0$, as $H = 0$ is not a center but corresponds to infinite values of τ and of the affine parameters along causal geodesics.

As mentioned in Part I, if $\rho' = 0$, the charge density q necessarily vanishes, hence UD solutions for which $\epsilon \neq 0$, labeled in the classification scheme of Table VII of Part I with Ch (i.e., ChUD solutions), are characterized by having an electric field given by

$$F_{tr} = -(-g^{tt})^{1/2} \epsilon c_0^3/r, \qquad (26)$$

where $F_{tr} = -F_{rt}$ is the only nonzero component of the electromagnetic Maxwell tensor $F_{\alpha\beta}$. It has been suggested by Mashhoon and Partovi,[12] in their study of gravitational collapse of spheres modeled on the ChMcV($r2$)(UD) solution, that these configurations describe neutral fluid being accreted by a charged black hole. Such an explanation, though, cannot hold in general (for example, if the fluid is not matched to a Reissner–Nordstrøm solution).

Since the electric field (26) is orthogonal to u^α, pointing in the direction $\partial/\partial r$ along the leaves of the surfaces Σ_T (see Sec. II), it is possible to infer the location of its sources by representing this field as arrows "combing" a surface Σ_T embedded in $\mathbb{R}^4$ by Eq. (4). As illustrated by Fig. 18(a), the zero net charge along the Σ_T in ChUD solutions is consistent to the wormhole topology of these surfaces, preventing the field lines converging into sinks and sources. The latter, following this model, must be charges of opposite signs located at the boundaries of $\mathscr{M}$ between which the Σ_T extend (the FD singularity, $r = 0$, $H = 0$, $\Pi = 0$, or in the direction $\partial/\partial r$ as $r \to \infty$), leading to an analogous situation to the elementary model of a dielectric in a two-plate capacitor [see Fig. 18(b)]. However, the wormhole topology of the Σ_T is not enough to account for the lack of net electric charge, as there are solutions with $q \neq 0$ having surfaces Σ_T with such topol-

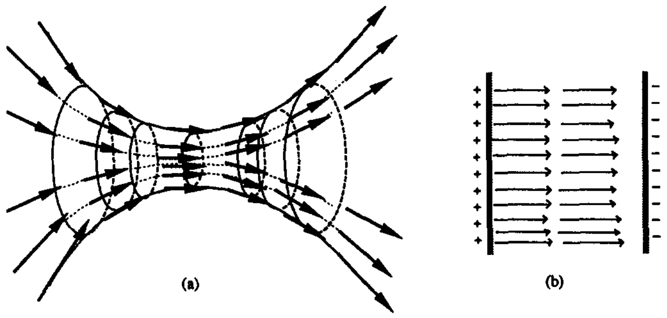


FIG. 18. Electric field in solutions without a center. For charged solutions in general, the electric field lines go along the surfaces Σ_T, as shown in (a) for the case when these surfaces have $S^2\times\mathbf{R}$ topology. This topology prevents the field lines from converging into sinks and sources, and so the electric field could be a polarization field analogous to that produced in the elementary dielectric capacitor of (b). The particular case of ChUD solutions (see Table VI of Part I), homeomorphic to $S^2\times\mathbf{R}^2$ and with zero net electric charge, corresponds to uniform dielectric polarization.

ogy (any charged solution without a center and with $\rho'\neq 0$). Following the model of a dielectric capacitor, the polarization charge vanishes if the polarization vector has a constant nonzero magnitude (uniform polarization, see Ref. 38). Hence, the neutral fluid in ChUD solutions could model a uniformly polarized dielectric medium, while the charged fluid in solutions without a center could model such a medium with nonuniform polarization.

The collapse of UD spheres in a Schwarzschild or Reissner–Nordstrøm background is a particular case of the situation discussed in Sec. VI, in which the interior fluid region is not compact and the null boundary $r=0$ behaves as a sort of white hole when the latter is singular (see Figs. 22–25). As mentioned in Sec. VI, the global view of these collapsing spheres has been overlooked by authors (McVittie,[10] Knutsen,[11] Mashhoon and Partovi,[12] and Glass[13]) studying the gravitational collapse of UD solutions.

IX. CONFORMALLY FLAT SOLUTIONS

Conformally flat ChKQ uniform-density solutions (CF solutions) are particular cases of the general conformally flat perfect fluid solution obtained by Stephani,[39] i.e., the "Stephani universe." The global structure of CF solutions has been studied by Cook[2] and Krasinski,[3] the latter author treating them as the spherically symmetric particular case of the Stephani universe. The remainder of this section aims to complement the work of these authors.

The metric coefficient H for CF solutions is given by Eq. I(48); however, since the form of the metric I(11) or II(1) with the values of $k=0,\ \pm 1$ denotes the same solution if $\rho'=0$ (see Appendix C), the function y in I(48) can be replaced by $r^2/2$ as in the last section (choice $k=0$). Using this choice, I(48) can be rewritten in a more convenient form as

$$H=H_c/(1+\tfrac{1}{4}LH_c^2r^2), \tag{27}$$

where the constant c_0 in I(48) has been set to unity and $H_c=H(t,0)=e^T/|L|$. As Eq. (27) indicates, conditions II(45) hold, and thus CF solutions (unlike UD solutions with $\Psi_{(2)}\neq 0$) have a regular center at $r=0$. The use of H_c as a t-parameter suggests using either one of the time coordinate choices II(31) or II(32), t being the proper time along the center $r=0$ in the latter. However, this coordinate choice will be left unspecified for the time being.

The t-parameters H_c and L in (27) relate to Eqs. (18) and (19) of Krasinski's paper by $L=k_{(\mathrm{Kr})}$ and $H_c=|R_{(\mathrm{Kr})}|$. The FRW solutions follow from (27) by setting $LH_c^2=k$. The metric coefficient $(-g_{tt})^{1/2}$ for these solutions is given by

$$\begin{aligned}(-g_{tt})^{1/2}&\\ &=U^{-1}\\ &=[(\dot H_c/H_c)/(\Theta/3)]\\ &\quad\times\frac{1-(r/2)^2LH_c^2[1+(\dot L/L)(\dot H_c/H_c)^{-1}]}{1+(r/2)^2LH_c^2},\end{aligned} \tag{28}$$

which imposes the restriction, connected with II(29) and analogous to (19), that $(\dot H_c/H_c)$ and $\dot L$ must be of the same order as Θ if expanded around t such that $\Theta(t)=0$. For H given by (27), Eq. II(16) becomes simply Eq. (23) with $c_0=1$, so that, following Krasinski,[3] $L(t)$ plays the role of a time-dependent "curvature index" k. Equations II(20) and II(28) are given by

$$\begin{aligned}&\frac{\Theta^2}{9}+\frac{d}{d\tau}\frac{\Theta}{3}+4\pi\left(p+\frac{\rho}{3}\right)\\ &\quad+(L/2)[1-(r/2)^2LH_c^2]\mathscr{B}=0,\end{aligned} \tag{29a}$$

$$\mathscr{A}=\frac{(r/2)LH_c^2}{1+(r/2)^2LH_c^2}\mathscr{B}, \tag{29b}$$

where

$$\mathscr{B}\equiv-\frac{2+(\dot L/L)(\dot H_c/H_c)^{-1}}{1-(r/2)^2LH_c^2[1+(\dot L/L)(\dot H_c/H_c)^{-1}]}. \tag{30}$$

A convenient (though not the only possible) choice of time coordinate is to take t as the proper time of comoving observers at $r=0$ [choice II(32)], leading to $\Theta/3=\dot H_c/H_c$. Thus if a localized equation of state is chosen at $r=0$ (see Sec. VII of Part II), one obtains $\rho=\rho(H_c)$ while the remaining t-parameter L can be used to impose a second localized equation of state at another surface Σ_r (say $r=r_0>0$). Given a pair of localized equations of state, the relation between the two t-parameters H_c and L is given by the constraint $\rho(H_c)=\rho(H_0)$, which follows from the Bianchi identity II(41) and Eq. (27). As an example, consider a sphere matched to a Schwarzschild exterior ($p_0=0$) such that $p_c=\rho/3$. In this case, the constraint relating L and H_c takes the form $\alpha H_c^{-4}=mH_0^{-3}$, where α is a constant and m is given by II(74c) with $J_0=0$, leading to

$$LH_c^2=(2/r_0^2)\,[(m/\alpha H_c)^{1/3}-1], \tag{31a}$$

so that

$$H=H_c^{4/3}/\{H_c^{1/3}[1-(r/r_0)^2]+(m/\alpha)^{1/3}(r/r_0)^2\}. \tag{31b}$$

The use of two localized equations of state has been the procedure followed by Thomson and Whitrow[40] and Bondi[41] in their study of CF solutions, though these authors chose a polytropic localized equation of state at the center of the sphere instead of $p_c = \rho/3$.

From the form of Eqs. (23), (29), and (30), the coordinate values satisfying

$$H = 0 \Leftrightarrow H_c = 0, \tag{32a}$$

$$\left.\begin{aligned} &(-g_{tt})^{1/2} = 0 \\ &\Theta \neq 0 \end{aligned}\right\}$$

$$\Rightarrow \mathscr{G}(t,r) = 1 - (r/2)^2 L H_c^2 [1 + (\dot{L}/L)(\Theta/3)^{-1}] = 0, \tag{32b}$$

mark two possible scalar curvature singularities analogous to II(48) and II(49). This can be easily verified, since the incompleteness criteria for causal curves given by II(52b) and the convergence of the integral in II(55) are satisfied as comoving observers and null geodesics approach these boundaries. The singularity marked by (32b) was mentioned by Krasinski,[3] but this author did not comment on its conformal structure and also ignored the singularity (32a), which can also be present in the more general Stephani solution[39] ($R = 0$ and $F = 0$ in Krasinski's notation). At $\mathscr{G} = 0$ in (32b), as $Q = 0$ in II(49), ρ remains finite, but p, p', and $\mathscr{A}$ diverge; hence this is also a "finite-density" (FD) singularity. On the other hand, at $H_c = 0$ in (32a) ρ, p, and p' diverge but $\mathscr{A}$ vanishes. The conformal structure of these singular boundaries can be found from the equation equivalent to II(56), which reads $(dt/dr)_{\rm null} = \pm (\Theta H_c/3)/\mathscr{G}$. Since $(dt/dr)_{\rm null} \to \pm\infty$ as $\mathscr{G} \to 0$, one has a spacelike FD singularity like that given by II(49) (see Fig. 2 of Part II).

However, the singularity marked by (32a) has a different structure from II(48). Since $H_c = 0$ coincides with a surface Σ_t, and $|(dt/dr)_{\rm null}| > 0$ as $H_c \to 0$, this singularity is spacelike (see Fig. 2 of Part II). As $\mathscr{A}$ and $a^\alpha{}_{;\alpha}$ in (29) remain finite, ρ and p diverge only if the equation of state (whether "localized" or not) is chosen so that $\Theta \to \pm\infty$ as $H_c \to 0$, though in CF solutions, $H_c = 0$ does coincide with $H = 0$, and so this singular boundary is a standard big bang (see Fig. 4 of Part II). In the particular case of the pair of localized equations of state ($p_c = \rho/3$, $p_0 = 0$), leading to Eqs. (31), $\mathscr{G} = 0$ in Eq. (32b) implies $H_c < 0$, and thus does not hold since the evolution of the fluid terminates at the big-bang singularity marked by $H_c(t) = 0$. However, if the choice of t-parameters allows for (32b) to occur, the evolution of the collapsing sphere will be qualitatively analogous to that outlined in Sec. XI of Part II for M- and W-type solutions in which $Q = 0$ occurs [types (i) and (ii)]. See Fig. 14(a) of Part II. Formation of the apparent horizon in CF collapsing spheres has been studied by Glass[13] and by Knutsen.[42]

As mentioned in the Introduction, the global topology of CF solutions was examined by Cook[2] and by Krasinski[3] in terms of the behavior of $R = fH$ and $\zeta = \int H\, dr$ along the surfaces Σ_t, as was done in previous sections for M- and W-type solutions. As mentioned in Sec. VIII, unlike UD solutions, the surfaces Σ_t in CF solutions do have constant curvature, in the sense of having a Ricci tensor of the form (2a). In fact, from (27), it follows immediately that the Σ_t are isometric to flat three-space $\mathbb{R}^3$ if $L(t) = 0$, to S^3 if $L(t) > 0$, and to $\mathbb{H}^3$ if $L < 0$.

As commented by Cook[2] and Krasinski,[3] the regularity boundary $\Pi = 0$ might occur, taking the form $\Pi = 1 + LH_c^2 r^2/4 = 0$, and its effect in the topology of the surfaces Σ_t is similar to that discussed in Sec. IV and Fig. 15 for W-type solutions with $k = 1$ in which $\Pi = 0$ occurs. From Eq. (27), for Σ_t surfaces along which $L > 0$, the function Π has no zeros, and the range of r along these surfaces Σ_t goes from zero to infinite. In this case one has $R \to 0$ and ζ finite as $r \to \infty$, indicating that these surfaces Σ_t have topology S^3 with $r \to \infty$ [case given by (5a)], marking an "antipodal" center to the one at $r = 0$. However, if $L < 0$, $\Pi = 0$ can occur for a finite value of r, say r_1, along a given surface Σ_t, then both R and ζ diverge as $r \to r_1$, indicating that these surfaces have topology $\mathbb{R}^3$. The value of t in which the topology of the surfaces Σ_t changes from S^3 to $\mathbb{R}^3$ is precisely $L(t) = 0$ and, as mentioned by Krasinski,[3] a solution in which L changes sign from positive to negative would have a global structure similar to the de Sitter space-time. The representation of the topological embedding (6) of the space-time manifold in the CF case in which L changes sign is qualitatively analogous to that of Figs. 10(a) and 10(b) (in the latter case, $H = 0$ occurring at finite τ and coinciding with a surface Σ_T). Hence, the space-time manifold has the same $S^3 \times \mathbb{R}$ topology as de Sitter space-time, though the local geometry is very different.

For the particular case of the CF(T) solution (see Table VII of Part I) discovered by Bonnor and Faulkes,[43] $L = $ const, and so there is only one t-parameter, implying that only one localized equation of state is possible. This latter solution is also a W-type solution, and all expressions derived and discussed in previous sections can be applied to it by setting $\epsilon = \mu = 0$ and $fh = $ const. Bonnor and Faulkes studied the case $L > 0$ as a model of an oscillating sphere matched to a Schwarzschild solution. In this case, the unbounded configuration is homeomorphic to $S^3 \times \mathbb{R}$, as in the cases $L = L(t)$ discussed above. However, if L is a negative constant, all the surfaces Σ_t have negative constant curvature and the space-time manifold is homeomorphic to $\mathbb{R}^4$ with embedding diagrams qualitatively similar to those of Figs. 10(c) and 10(d).

For whatever choice of equation of state allowing $\Pi = 0$ to occur, from Eqs. II(7), (29), and (30), as $\Pi \to 0$ along the world lines of comoving observers, $L < 0$, $H_c > 0$, and Θ tend to constant values, and so one has the "asymptotically de Sitter" behavior characterized by $H \to \exp[\Theta\tau/3]$, and by ρ, p taking the asymptotical values similar to Eqs. (7) which, as mentioned by Cook[2] and Krasinski,[3] lead to physically unacceptable negative pressures. However, if $L > 0$ for all times, negative pressures might be avoided for all times.

The fact that CF solutions can be related to FRW and de Sitter solutions by conformal transformations is a strong indication that their conformal structure might be qualitatively similar to that of the latter solutions. Since $H_c = 0$ is a standard big bang then, except for the possible occurrence of a singularity at $\mathscr{G} = 0$, if L is everywhere positive, one has qualitatively the same global features as in a $k=1$ FRW solution. For the case when L is negative [or becomes negative if $L=L(t)$], the boundary $\Pi=0$ is spacelike. This follows from inserting $\Pi=0$ into the equation $(dt/dr)_{\rm null}$, leading to $(dt/dr)_{\rm null}\to\pm\infty$, and so this boundary is a spacelike $\mathscr{I}_{\pm}$ surface similar to that found in a de Sitter space-time[8] (see Fig. 11). Therefore, depending on the choice of equation of state (i.e., choice of t-parameters), the Penrose diagram of a CF solution in which $\Pi=0$ occurs could have some features qualitatively similar to the Penrose diagram of a FRW solution (for example a big bang) and, near $\Pi=0$, the spacelike null infinity of a de Sitter solution. Some possible Penrose diagrams for CF solutions are displayed in Fig. 19.

Conformally flat solutions for which $L>0$ and $|\mathscr{G}|>0$ hold everywhere are probably the only ChKQ solutions which have physically appealing local and global properties. In Secs. XI and XII, these solutions will be suggested as kinetic theory models of collisionless gas mixtures and local inhomogeneities in a cosmological background.

X. STATIC LIMITS

As mentioned in Sec. V of Part I, if the t parameters T and L are set to be constants, there is no time dependence of the metric coefficient H. The class of static spherically symmetric perfect fluid solutions obtained by setting $H=H(r)$ are characterized by the metric

$$ds^2=-G^2(r)dt^2+H^2(r)\times[dr^2+f^2(d\theta^2+\sin^2\theta\,d\phi^2)],\qquad(33)$$

where, given $H(r)$, the metric coefficient $G(r)$ can be obtained from the Einstein (or Einstein–Maxwell) field equations arranged in the form $(G^r{}_r-G^\theta{}_\theta)=8\pi(T^r{}_r-T^\theta{}_\theta)$ ("equation of pressure isotropy," see Ref. 44). This constraint can be expressed as

$$\frac{G_{yy}}{G}+\frac{2Y_y}{Y}\frac{G_y}{G}+3\mu h^5Y-\epsilon^2h^6Y^2=0,\qquad(34)$$

where $Y\equiv H^{-1}$ and $G_y\equiv dG/dy=fG'$. It is sufficient to know $H(r)$ in order to find out the matter-energy density $\rho=\rho(r)$ associated with a given static solution with metric (33). This form of ρ can be computed from the field equation $G^t{}_t=8\pi T^t{}_t$ obtained from (33), leading to

$$\frac{8}{3}\pi\rho(r)=\frac{1}{6}\,{}^{(3)}\mathscr{R}=\frac{1}{R^2}\left[1-\frac{2J}{R}+\frac{E^2}{R^2}-\left[\frac{(fh)'}{h}\pm(fh)^2\mathrm{Q}^{1/2}\right]^2\right],\qquad(35)$$

which, as expected, coincides with Eqs. II(16b) with $\Theta=0$. Therefore, the problem of finding G from a given H is equivalent to finding the four-acceleration (and thus p and p') which will keep the matter energy distribution associated with (35) in a static frame. The pressure p can be calculated from the Raychaudhuri equation:

$$4\pi\left(p+\frac{\rho}{3}\right)=a^\alpha{}_{;\alpha}=\Psi_{(2)}+\left[\frac{(fh)'}{h}-(fh)^2\mathrm{Q}^{1/2}\right]\frac{fG'}{G},\qquad(36)$$

while p' follows from II(22) with $\mathscr{A}=G'/G$.

It can be easily verified that Eq. (34) is satisfied by the following expressions for $G(r)$:

$$G(r)=\mathrm{Q}^{1/2},\qquad(37a)$$

$$G(r)=\mathrm{Q}[b_1\mathscr{N}(r)-b_2],\qquad(37b)$$

where Q is the same expression of Eq. II(30) with $H=H(r)$, b_1 and b_2 are arbitrary constants, and $\mathscr{N}=\mathscr{N}(r)$ must satisfy $\mathscr{N}'=\mathrm{Q}^{-1/2}$, which is the same con-

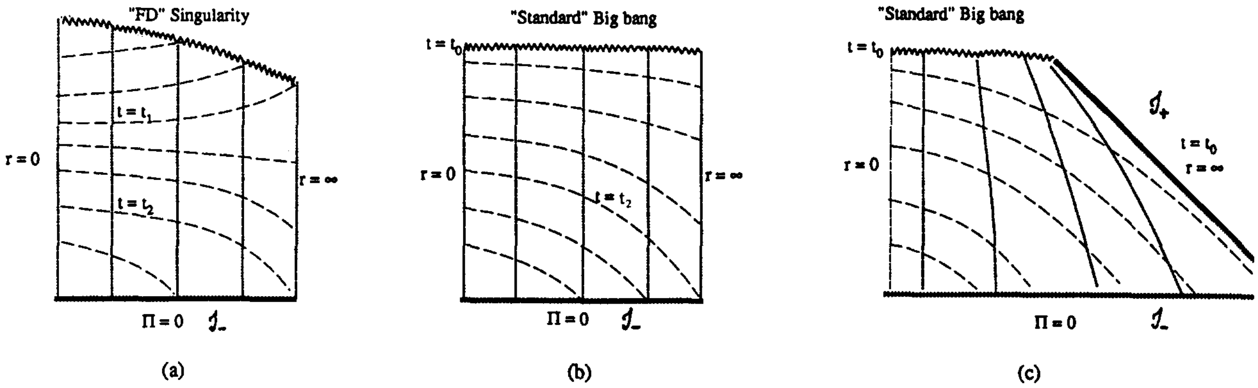


FIG. 19. Qualitative Penrose diagrams for some conformally flat solutions. Surfaces Σ_t and Σ_r are depicted as horizontal and vertical curves, while boundaries at infinity (i.e., $\mathscr{I}_{\pm}$ surfaces) are marked by thick solid lines. In (a) and (b), the t-parameter $L(t)$ changes sign from negative to positive at $t=t_2$, so that the $\mathbf{R}^3$ topology of the surfaces Σ_t reaching $\Pi=0$ changes to S^3. In (a), causal curves terminate at the FD singularity $\mathscr{G}(t,r)=0$, while in (b) these curves terminate at the standard big bang $H_c=0$ which coincides with $t=t_0$. (c) corresponds to the case when $L(t)$ is everywhere negative, so that all the Σ_t reach $\Pi=0$ and are isometric to the 3-D "pseudosphere" $\mathbf{H}^3$. The space-time manifold in the cases (a) and (b) is homeomorphic to $S^3\times\mathbf{R}$, while in (c) it has $\mathbf{R}^4$ topology. The existence of a null $\mathscr{I}$ surface in (c) can be inferred from the arguments illustrated in Fig. 25 (see Appendix A).

dition [Eq. (17)] that $\mathcal{N}(t,r)$ has to satisfy in the nonstatic case with $L = L(t)$. If $\epsilon = \mu = 0$, the solution of Eq. (34) with $y = r^2/2$ is given by

$$G(r) = a_1 + a_2[1 + LR^2(r)]^{1/2}, \tag{37c}$$

where a_1, a_2, and L are arbitrary constants and $R = rH$, with H obtained from (27) by setting L and H_c as constants. As expected, these expressions coincide with those which could be obtained from the forms of $G(t,r)$ for nonstatic solutions [i.e., Eqs. II(31) to II(33), (15), and (28)] by demanding that the limit of $(\dot{H}/H)/(\Theta/3)$ be finite as $\Theta \to 0$. Therefore, the static limits of ChKQ solutions are those solutions with metric (33) in which $G(r)$ has the forms (37a) for M- and W-type solutions, (37b) for solutions with $L = L(t)$ and $\Psi_{(2)} \neq 0$ (see Sec. VIII), and (37c) for CF solutions (see Sec. IX).

The static limits of ChKQ solutions will be regular (in the sense of Sec. VIII of Part II) if the constant parameters are chosen so that the metric coefficients $H(r)$ and $G(r)$ are non-negative functions for all the range of values of r. This choice is equivalent to the selection of a Σ_t surface of a given nonstatic solution, preferably a Σ_t along which (4-D) curvature scalars are bounded, and then a regular static ChKQ space-time is formally defined as the product $R \times \Sigma_t$. Thus, since the Killing vector $\partial/\partial t$ generates a one-parameter group of motions, all hypersurfaces Σ_t (i.e., the static frames) are isometric to each other. Provided that *ad hoc* topological identifications are avoided, static ChKQ solutions will belong to the same homeomorphic class of equivalence as their nonstatic analogs. However, there might be nonstatic solutions whose static limits are necessarily singular.

The metric (33) can be expressed in terms of R (curvature coordinates) by using $dR = R'\,dr$ and Eq. II(24), leading to

$$ds^2 = -G^2(R)dt^2 + \frac{dR^2}{[(fh)'/h \pm (fh)^2 Q^{1/2}]^2} + R^2(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{38}$$

where now f, h, and fh given by II(2), I(17), and I(23) are functions of R. If $fh \neq$const, one has $\rho' \neq 0$, and the metrics (33) or (38) describe spherically symmetric, static, perfect fluid solutions characterized by (35)–(37). It is quite possible that some of these static solutions are already known, for example, Glass and Mashhoon[16] and Mashhoon and Partovi[12] discovered that the static limits of the ChMcV$(r2)(X1,2)$ and NMcV$(r2)(X1,2)$ solutions (see Part I) correspond to a static solution (and its charged version) derived by Buchdahl[15] as a relativistic generalization of Newtonian polytropes of index 5.

As with their nonstatic analogs, the above-mentioned static solutions can be matched to Schwarzschild or Reissner–Nordstrøm space-times. Matchings of this type (though not restricted to static limits of ChKQ solutions) have been considered by Künzle[45] for regular spheres without a center, so that the whole space-time has $S^2 \times \mathbb{R}^2$ topology. Künzle found that large pressures ($|p| \gg \rho$) must occur inside of these spheres, though as shown by Eqs. (14), this does not necessarily occur in nonstatic ChKQ spheres with this topology.

The static limits of uniform density solutions are easier to appreciate from (38). If $(fh)' = 0$ and $G(r)$ is given by (37a) [or (24b)] with $L = 0$ (i.e., static limit of uniform density M-type solutions), $p = \rho = 0$, and the metric (38) reduces to that of Reissner–Nordstrøm solution without the Eddington–Finkelstein or Kruskal extensions [Eq. II(72) with $1 - 2m/R + e^2/R^2 > 0$, with $m = \mu c_0^5$ and $e^2 = \epsilon^2 c_0^6$]. Thus, for nonstatic UD M-type solutions of type (ii), the regularity condition $Q > 0$ [Eq. II(47)] reduces to $1 - 2m/R + e^2/R^2 > 0$ in the static limit. If $(fh)' = 0$ and $G(R)$ is given by (24a) with $L \neq 0$ (static limit of uniform-density W-type solutions), $8\pi p = -3c_0^2 L = -8\pi\rho$, and (38) becomes the metric of the Reissner–Nordstrøm solution with a cosmological constant $c_0^2 L$. Notice that the static limit for solutions with $\mu < 0$ [type (iv) if $\epsilon = 0$] leads to $m < 0$. For CF solutions, the metric (38) with $(fh)' = 0$ and $G(r)$ given by (37c), leads to the interior Schwarzschild solution[44] if $a_1 \neq 0$, and to the de Sitter solution if $a_1 = 0$. Therefore, the static limits of UD solutions will have (in general) the same $S^2 \times \mathbb{R}^2$ topology (if $\Psi_{(2)} \neq 0$) and $S^3 \times \mathbb{R}$ (if $\Psi_{(2)} = 0$) structure as their nonstatic analogs. By this criterion, the static limit for UD solutions with $\Psi_{(2)} \neq 0$ and $L = L(t)$ [metric (38) with $(fh)' = 0$ and G given by (37b)] should be Schwarzschild or Reissner–Nordstrøm (if $\epsilon \neq 0$) solutions with a "cosmological constant" related to L, as the latter solutions also have $S^2 \times \mathbb{R}^2$ topology.

Newtonian limits of Buchdahl's solution,[15] mentioned previously, were discussed by Glass and Mashhoon[16] and Mashhoon and Partovi.[12] Newtonian limits for other ChKQ solutions should follow by setting (33) into the "weak field" approximation $G(r) \approx 1 + 2\phi(r)$, $H \approx 1 - 2\phi(r)$ with $\phi \ll 1$ identifying the Newtonian potential (Sec. 17.4 of Ref. 32). From the forms of H derived in Part I, it seems that such a weak field approximation can be realized for M-type solutions, resulting in $\phi(r)$ somehow related to the function $X(r)$ defined in I(24b). However, for W-type solutions, the term with the factor L in (37a) seems to correspond to a repulsive force associated with a cosmological constant. Hence it is not clear if a Newtonian limit would be possible in this case.

XI. KINETIC THEORY MODELS

As an alternative to supplying an equation of state based on the criteria discussed in Secs. VI and VII of Part II, the time evolution of a given ChKQ solution could be related to the physical properties of the fluid via relativistic kinetic theory. Although a full kinetic theory treatment of these solutions is not attempted in this section, useful guidelines for such treatment follows from previous research in this topic,[27,46–48] and the properties of ChKQ solutions presented hitherto in this paper and in Part II. The discussion of this section will be restricted to neutral solutions.

From the point of view of kinetic theory, one assumes that the fluid is modeling a mixture of gases associated with a non-negative distribution function $\mathcal{F}(x_N{}^\alpha, p_N{}^\alpha)$, where $x_N{}^\alpha$, and $p_N{}^\alpha$ are the coordinates and momenta of the Nth component of the mixture. Knowing $\mathscr{F}$, the energy momentum tensor (and so macroscopic state variables such as p and ρ) follows from evaluating[48]

$$T^{\alpha\beta} = \sum_N \int_\pi p_\alpha^N p_\beta^N \mathscr{F}\, d\pi. \qquad (39)$$

Other state variables such as N^α and S defined in Sec. VI of Part II follow from similar integrals of $\mathscr{F}$ (see Ref. 48), and so the field equations become a determined system, i.e., the Einstein–Boltzmann equations. However, for any functional form of $\mathscr{F}$, ρ, and p obtained from (39) must satisfy the strong energy condition $p+\rho/3 \geqslant 0$, which for nonstatic M- and W-type solutions does not hold throughout the evolution of the fluid because the curvature terms ${}^{(3)}\mathscr{R}$ and $a^\alpha_{;\alpha}$ diverge at a different rate near $H=0$ and/or $Q=0$, and so $|p| \gg \rho$ at some point in the evolution of the fluid (see Sec. X of Part II and Secs. IV and V). On the other hand, for regular static M- and W-type solutions and CF solutions with $L>0$ and $|\mathscr{G}|>0$, the lack of singularities at which $|p| \gg \rho$ (see Secs. IX and X) means that the conditions on curvature terms ${}^{(3)}\mathscr{R}$ and $a^\alpha_{;\alpha}$ allowing for $p+\rho/3 \geqslant 0$ to hold in the Raychaudhuri equations [(29) and (36)] are not straightforwardly inconsistent, and so could hold in either case for a "good" choice of parameters. Hence, among all ChKQ solutions, only static and CF solutions could be compatible with a kinetic theory approach. In either case, an equation of state connected with physical assumptions on the microscopic components of the fluid would appear from the distribution function. Such an equation of state cannot be barotropic (see Sec. VI and Appendix A in Part II), and so p will depend on ρ and the entropy density $S(r)$. However, under a kinetic theory approach, the dependence of state variables on position along the surfaces Σ_t introduced by $S(r)$ could reflect a physically reasonable situation, such as the position dependence of the concentration of different components of the gas mixture.

For a perfect fluid energy-momentum tensor, the entropy production must vanish and so from the relativistic H theorem (Ref. 48), $\mathscr{F}$ describes either a collisionless gas mixture or a gas mixture in which collision integrals add up so that entropy production vanishes (this effect is called "detailed balancing"). In either case, $\mathscr{F}$ must satisfy the Liouville equation

$$\left[p^\alpha \frac{\partial}{\partial x^\alpha} - \Gamma^\alpha_{\beta\gamma} p^\beta p^\gamma \frac{\partial}{\partial p^\alpha}\right] \mathscr{F} = 0, \qquad (40)$$

where, for the sake of simplicity, a one-component gas mixture has been considered (the generalization to N components is immediate). For the case of detailed balancing, the Liouville equation (40) leads to the so-called "isotropic" distributions.[46–48] For such distributions, the fluid is restricted to be shear-free, nonrotating, or nonexpanding (i.e., $\omega\Theta = 0$, where ω is the magnitude of the vorticity vector). Although an isotropic $\mathscr{F}$ always leads to a perfect fluid energy-momentum tensor via (39), the converse is not true.[49] The relativistic version of the Maxwell–Boltzmann distribution known as the Jüttner distribution[27,48,50] is an important particular case of isotropic distributions given for one-component mixtures by $\mathscr{F} = \exp[\mathscr{b}(x^\alpha) + \lambda_\beta(x^\alpha)p^\beta]$, where λ^β is a timelike Killing vector if $m>0$ and a conformal timelike Killing vector if $m=0$. Combining (40) with (39) leads to $\lambda^\alpha = \mathscr{T} u^\alpha$, where $\mathscr{T}$ is the macroscopic temperature in II(37) (Ref. 51). Apparently, λ^α could be identified with the vector λ^α defined by Eqs. II(10) and II(11). However, the latter is not a conformal Killing vector, and as discussed in Sec. VI of Part II, Eq. II(37) is inconsistent with condition II(39b) in nonstatic ChKQ solutions and so the identification of $\mathscr{T}$ with $(-g^{tt})^{1/2}$ as a sort of nonstatic analog of Tolman's law[52] [Eq. II(40)] will not hold for these solutions. On the other hand, Eq. II(37) is not inconsistent with II(39b) for static ChKQ solutions, and so if the strong energy condition is satisfied, these solutions could model a detailed balancing gas mixture with $m>0$ associated with a Jüttner distribution function. This would also be in agreement with the fact that in this case the coefficient of thermal conductivity κ need not vanish.[50,51,53]

For collisionless gas mixtures, a method for solving Eq. (40) has been developed by Ehlers[48] on the basis of the existence of Killing vectors. This method has been usually applied to FRW and spherically symmetric static solutions by Fackerell,[17] Ray and Zimmerman,[18] and Ray[19] (see Maharaj[20] for a review of previous research on collisionless gases). Obviously, static ChKQ solutions compatible with the strong energy condition can be thought as examples of these collisionless models, and this aspect has been recalled by Glass and Mashhoon[16] and Mashhoon and Partovi[12] in connection with the static limits of the particular solutions ChMcV($r2$)(X1,2) and NMcV($r2$)(X1,2). However, a collisionless gas model might be also applicable to nonstatic CF solutions complying with the strong energy condition. For these solutions (as with all other nonstatic ChKQ solutions), II(39b) is inconsistent with II(37) implying that $\kappa = 0$. Intuitively, this situation seems to be incompatible with the existence of collisions, and from kinetic theory calculations of this coefficient,[50,53] it can be appreciated that, indeed, κ vanishes if there are no collisions. Hence, CF solutions could model such a gas mixture, and this could be tested following the study of Petrov G_4 VIII cosmological models carried on by Ray and Zimmerman[18] and Ray.[19] Such a kinetic theory treatment of CF solutions will be attempted in a future paper.

XII. COSMOLOGICAL APPLICATIONS (NEUTRAL SOLUTIONS ONLY)

As mentioned in Sec. II of Part II, SSSF solutions are characterized by an isotropic rate of change of relative distances of neighboring particles along the surfaces Σ_t (rest frames of comoving observers). However, in solutions other than FRW, this kinematic isotropy, characterized by the one-parameter group of conformal motions generated by the vector field II(10), is not actually "detected" by the comoving observers in the form of an isotropic Hubble law. This is so because of the radial dependence of the four-velocity $u^\alpha = U(t,r)\delta^\alpha_t$, which introduces a pair of preferential directions along the four acceleration vector a^α defined by II(9a). Hence, as shown by Fig. 20, an arbitrary comoving observer at $r>0$ (assuming that $r=0$ marks a regular center) detects a position- and direction-dependent red-shift distribution of photons emitted from neighboring observers. An expression for this red-shift distribution has been derived by Ellis[54] and Collins[55]:

$$\frac{dz}{z} = \left[\frac{\Theta}{3} + a_\alpha x^\alpha\right] dl, \tag{41a}$$

where dl is the magnitude of the connecting vector $X^\alpha = h^\alpha_\beta x^\beta\, dl$ along the rest frames, and $x_\beta x^\beta = 1$. Equation (41a) shows the combined effect of a "monopole" term $\Theta/3$ and a "dipole" term due to the acceleration a^α. For FRW solutions, only the monopole contribution survives, and one has an isotropic red-shift distribution. For SSSF solutions other than FRW, both dipolar and monopolar contributions shape this red- (blue-) shift distribution. Selecting an arbitrary comoving observer labeled by the coordinates $(r_0 > 0, \theta = \pi/2, \phi_0 + \delta\phi)$ (see Fig. 20), and assuming that it receives photons emitted by neighboring observers at the same proper distance along the rest frame given by an arbitrary surface Σ_t, Eq. (41a) becomes

$$\frac{dz}{z} = \left[\frac{\Theta}{3} + \mathscr{A} \cos \psi\right] dl, \tag{41b}$$

where ψ is a "telescopic" angle such that $\psi = 0$ along the direction of the four acceleration (see Fig. 20). Equation (41b) indicates a maximum red shift/blue shift in the pair of privileged directions parallel to the four-acceleration. If this arbitrary observer approaches a FD singularity given by II(49) or (28b), $\mathscr{A}$ diverges while Θ remains finite, and so this observer would detect infinite red/blue shifts in the pair of directions parallel to $a^\alpha(\psi = 0,\pi)$. This situation would also occur in the AD big bang at $H = 0$ in the infinite past (future) of comoving observers in type (i) or (iv) ChKQ solutions with $\Psi_{(2)} \neq 0$ [Eq. II(48)].

However, if the arbitrary receiving observer comoves along a regular center (whether such a center is labeled by $r = 0$, $r = \pi$, or $r = \infty$), since conditions II(45) hold there, $\mathscr{A}$ vanishes and only the monopole term $\Theta/3$ survives in Eqs. (41). This leads to an isotropic Hubble law, connected to the fact that such centers are fixed points of SO(3), and thus space-time is isotropic with respect to these loci, though, as commented in Sec. I of Part II, the same situation would occur in such symmetry points in any spherically symmetric space-time. If there is more than one regular center, as in solutions with $S^3 \times \mathbb{R}$ topology, one has $\mathscr{A} = 0$ along both of them, but the local geometry (curvature, density, pressure, etc.) in one regular center is in general different from that at the other. In solutions with topology $S^2 \times \mathbb{R}^2$ (see Secs. V, VI, and VIII), there are no regular centers and so there are no privileged comoving observers detecting an isotropical Hubble law. Equations (41) are local equations, and so must be integrated in order to have more information about the red-shift distribution of distant observers and the microwave background.

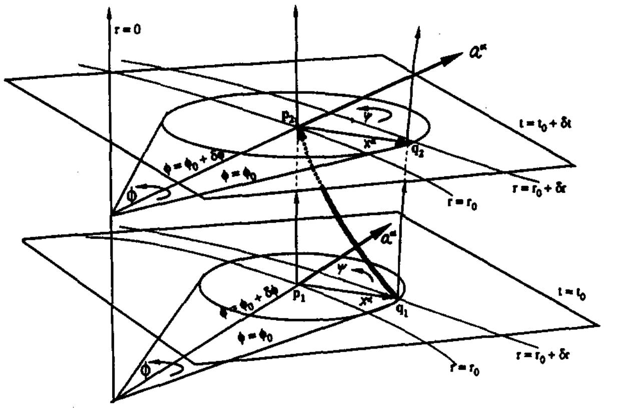


FIG. 20. Emission and reception of photons between neighboring comoving observers. A photon (thick gray arrow) is emitted at $t = t_0$ by the observer with world line q_1q_2, whose spatial coordinates are $r_0 + \delta r$, $\theta = \pi/2$, and ϕ_0. This photon is received at $t = t_0 + \delta t$ by the neighboring observer p_1p_2 whose spacial coordinates are r_0, $\theta = \pi/2$, and ϕ_0. The red-shift distribution for similar photons [Eq. (41)] contains a term which depends on the relative direction between the "connecting" vector X^α (p_1q_1 and p_2q_2) and the four-acceleration (thick arrows). This dipolar term can be expressed in terms of the cosine of the "telescopic" angle Ψ. For the observer comoving along the center $r = 0$, the latter dipolar term vanishes, and so this privileged observer detects an isotropic Hubble law.

From the discussion of previous sections, the only type of ChKQ solution that could provide a relatively less physically objectionable cosmological model would be a CF solution without the FD singularity ($|\mathscr{G}(t,r)| > 0$) and with $L(t)$ positive for all t (see Sec. VIII). Such a cosmological model would share the basic global features characterizing a $k = 1$ FRW solution (a standard spacelike big bang, $S^3 \times \mathbb{R}$ topology), but would have very different local properties as a result of the existence of a pressure gradient associated with the four-acceleration (29b). Although this position-dependent pressure could somehow be justified physically (see Sec. XI), such a spherically symmetric cosmological model would violate the Copernican principle by having observers comoving at the regular center $r = 0$ as a class of privileged observers. On the other hand, solutions homeomorphic to $S^2 \times \mathbb{R}^2$ (such as McVittie's[9] with $k \neq 1$, see Appendix B), lacking such a privileged symmetry center, are not incompatible with the Copernican principle, although they might be objectionable on the grounds of their local properties (energy conditions do not hold) and/or because of the presence of unphysical singularities (the FD singularity, or the possibility of having a singular null $\mathscr{I}$, see Sec. V). Hence it seems that the only physically acceptable cosmological application of neutral ChKQ solutions is in the form of a model of a spatially localized, spherically symmetric cosmological inhomogeneity immersed in a FRW background.

The use of ChKQ solutions as models of cosmological inhomogeneities has been suggested previously by Eisenstaedt.[14] This author considered a CF solution as a local inhomogeneity (his region I) matched to a cosmological background consisting of a UD solution with $\Psi_{(2)} \neq 0$ and $L = L(t)$ (his region II), while in a second paper he constructed a "coating model" by matching to the CF solution a series of similar UD solutions. However, Eisenstaedt did not examine the singularity structure and global view of these hybrid configurations, all constructed with solutions discussed in Secs. VIII and IX. Since Eisenstaedt's exterior region does not cover $r = 0$, the null $\mathscr{I}$ at this locus does not occur. This boundary would be necessarily singular as these

UD solutions have $L \neq 0$ and so Q vanishes as $r \to 0$ (see Table I and Secs. V and VIII). If the exterior UD solution is a type (ii) solution and/or the CF interior solution allows for $\mathscr{G} = 0$ to happen [Eq. (32b)], the spacelike FD singularity would arise as $(-g_{tt})^{1/2}$ given by (15) and (28) vanishes. If the exterior solution is of type (iv), then it must be matched to a CF solution with $|\mathscr{G}| > 0$. The singularity structure in the latter case would consist of a standard big bang as $H_c \to 0$ in the CF solution [Eq. (32a)], while in the UD solution one can have either a combination of L and FV singularities together with a timelike AD big-bang regular boundary (see Sec. VIII). In the best case, it might be possible to set the parameters of the UD solution as in Eqs. (20) and (21) and Fig. 17(a), one would have a combination of standard and AD big bangs joined by a L singularity. In all cases, the "asymptotically de Sitter" behavior associated with Eqs. (6) and (7) would arise if $L(t)$ changes sign in the CF solution or if the parameters of the UD solution allow for $\Pi(t,r) = 0$ to happen. Although the CF solution might have reasonable properties (see Secs. IX and XI), as discussed in Secs. IV, V, and VIII, the asymptotical features of UD solutions with $\Psi_{(2)} \neq 0$ are physically unappealing, and so Eisenstaedt cosmological configurations are theoretically interesting but unacceptable from a cosmological point of view.

A more suitable model of a spherically symmetric shear-free perfect fluid inhomogeneity could be constructed by matching a CF solution, with $\rho = \rho_{\rm int}$ and metric given by (27) and (28), directly to a $k = 1$ FRW background with $\rho = \rho_{\rm ext}$ and metric

$$ds^2 = -dt_*^2 + \{H^2(t_*)/[1 + r^2/4]^2\} \times [dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)] \tag{42}$$

along an arbitrary surface Σ_r labeled by $r = r_0 > 0$ (the "matching surface"). If the time coordinate in the CF region is chosen as the proper time of observers comoving with the matching surface, then this coordinate coincides with the cosmic time t_* at $r = r_0$ and so, the matching surface can be parametrized by the coordinates $y^a = (t,\theta,\phi)$. However, such a matching would not be continuous (a "C^0 matching") following Darmois matching conditions.[56] The latter require the first and second curvature forms

$$\mathscr{G}_{ab} \equiv J^\alpha_a J^\beta_b g_{\alpha\beta}, \tag{43a}$$

$$K_{ab} \equiv -n_{\alpha;a} J^\alpha_b + \Gamma^\alpha_{\beta\gamma} J^\beta_a J^\gamma_b n_\alpha, \tag{43b}$$

where

$$J^\alpha_a \equiv \frac{\partial x^\alpha}{\partial y^a}, \quad x^\alpha \equiv (t, r_0, \theta, \phi),$$

and n^α is a unit vector normal to the matching surface, to coincide for both space-times at this three-surface. Although it is always possible to have the first curvature form (43a) [i.e., the metric (42) coinciding with the metric II(1) at $r = r_0$], the second fundamental form (43b) (the Gaussian or extrinsic curvature of this Σ_r surface) will not agree at the matching surface. This is connected to the fact that the pressure gradient (29a) cannot be made to vanish at an arbitrary Σ_r surface $r = r_0 > 0$, and at the same time having the first curvature form (i.e., the metric) coinciding with (42) at r_0. In some solutions with $L = L(t)$ but $\Psi_{(2)} \neq 0$, the second t-parameter $L(t)$ could be set in such a way that both curvature forms (43a) and (43b) coincide with those of a FRW solution at the matching surface. However, these solutions, besides having complicated metric coefficients, might present unphysical singularities (FD singularity or AD big bang), and so are likely to be unsuitable as models of local inhomogeneities. Therefore, the only possibility left is a "Swiss cheese" model in which a CF solution with $L(t) > 0$ and $|\mathscr{G}(t,r)| > 0$ is matched to a Schwarzschild "vacuole," which in turn matches to a FRW background, not necessarily with $k = 1$. Such a model of a local inhomogeneity could be related to matter condensation processes, black-hole formation or to early universe phenomena via a kinetic theory study of CF solutions, as proposed at the end of Sec. XI.

ACKNOWLEDGMENTS

Again, I wish to thank Professor M. A. H. McCallum and Professor C. J. S. Clarke for their interest in this work and for their valuable and illuminating comments and suggestions.

I am indebted to the National University of Mexico (UNAM) for providing financial support.

APPENDIX A: THE WYMAN SOLUTION, GLOBAL VIEW

The Wyman solution has been studied from a global point of view by Mashhoon and Partovi[5] and by Collins.[6] This Appendix will complement such a study by applying to this solution the results developed in Secs. II–IV. As in Appendix A of Part II, only the neutral Wyman solution will be considered. In order to relate the discussion of this Appendix to that of Collins' paper, it is useful to look at Appendix A and Table II of Part II.

From Table II of Part II, the boundaries $Q = 0$, $H = 0$, and $\Pi = 0$ are given in the (T,r) representation [coordinate choice II(33)] as parabolas of the type $T + r^2/2 = \alpha_0$, where α_0 is a constant depending on the parameters of the solution. In their study of the Wyman solution, Collins[6] and Mashhoon–Partovi[5] have used the variable $v = T + r^2/2$, which can be characterized invariantly as labeling hypersurfaces of constant Hubble scale factor [i.e., $H(T,r) = \text{const}$]. These hypersurfaces appear in a (T,r) coordinate diagram (see Fig. 1 of Part II) as parabolas corresponding to different values of the constant α_0. However, unlike the surfaces Σ_T and Σ_r, which can be globally characterized as spacelike and timelike hypersurfaces, the hypersurfaces $v = \text{const}$ do not have such a globally defined conformal structure, changing in general from being locally timelike to being locally null and spacelike. Hence, their usefulness in a global study of the Wyman solution (or any other ChKQ solution) seems to be limited.

Also, the coordinate choice II(33), used by Mashhoon–Partovi and Collins, breaks down (i.e., it has a coordinate singularity) at the surface Σ_T such that $\Theta(T) = 0$ (see Sec. V of Part II). Therefore, it is convenient to study the Wyman solution in terms of the time coordinate t defined by II(31) which satisfies $dT/dt = -\Theta/3$. Using Eqs. II(A2), the coordinate transformation relating t and T is

$$t(T) = \int \frac{dT}{[L_0 - 6LT]^{1/2}} = -\frac{1}{3L}[L_0 - 6LT]^{1/2} \tag{A1a}$$

$$\Rightarrow T(t) = L_0/6L - (3L/2)t^2, \tag{A1b}$$

so that the expansion, as a function of t (but no longer appearing in the metric), is given for the case $L \neq 0$ [Eq. II(A2a) by

$$\Theta/3 = -3Lt, \tag{A1c}$$

implying that $\Theta = 0$ coincides with $t = 0$. For the case $L = 0$ [Eq. II(A1b)], Θ is a constant and one can simply set $T(t) = t$ without loss of generality.

The Wyman solution in all cases classified in Table II of Part II corresponds to the three-parameter $\{\mu, L, L_0\}$ family of space-time manifolds $\mathscr{M}$ homeomorphic to $\mathbb{R}^4$, whose metric is now II(31), with $f(r) = r$ $(k = 0)$, $Q = -2\mu H^{-1} + LH^2$, and with $H(T(t), r)$ given explicitly in terms of (t,r) by Eqs. I(36) and (A1a). The domain of regularity in these coordinates corresponding to various particular cases are shown in Fig. 21. The embedding diagrams, for the cases corresponding to types (i), (ii), and (iv), are qualitatively analogous to those shown in Figs. 7(c), 10(c), and 10(d), respectively. These types will be considered separately. As will be shown below, and in agreement with Mashhoon and Partovi and with Collins, the various particular cases of the Wyman solution have physically unappealing global properties. As far as their global view is concerned, the Wyman solution is qualitatively analogous to other similar W-type solutions incompatible with a barotropic equation of state (see Appendix A of Part II).

1. Type (ii)

In this case, the boundaries of $\mathscr{M}$ are the FD singularity $Q = 0$ and the asymptotically de Sitter boundary $\Pi = 0$. In the (t,r) representation, from Table II of Part II and Eqs. I(36) and (A1), these boundaries are marked by the following constraints:

$$Q(t,r) = 0 \Rightarrow t^2 - r^2/3L = L_0/9L^2, \tag{A2a}$$

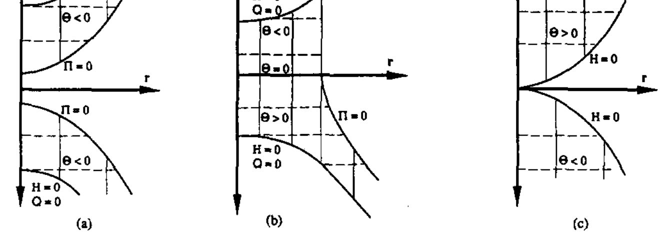


FIG. 21. Domain of regularity for the various cases of the Wyman solution. (a) corresponds to the case, whether type (ii) or (iv) ($L \neq 0$), in which the fluid expands ($\Theta > 0$) or collapses ($\Theta < 0$) monotonously. (b) depicts the domain of regularity for a configuration in which the fluid layers labeled by $0 < r < r_1$ bounce from $\Theta > 0$ to $\Theta < 0$ at $t = 0$. The case type (iv), with $L = 0$ and Θ constant (positive or negative), is considered in (c).

$$\Pi(t,r) = 0 \Rightarrow t^2 - \frac{r^2}{3L} = \frac{1}{9L}\left[\frac{L_0}{L} - \alpha_0\right], \tag{A2b}$$

$$\alpha_0 \equiv \mathrm{cn}^{-1}[(\sqrt{3} - 1)/(\sqrt{3} + 1)]. \tag{A2c}$$

The FD singularity marked by (A2a) is everywhere spacelike (see Fig. 2 of Part II), however, the conformal nature of $\Pi = 0$ must be determined from condition (10) by comparing the slopes of the null cones near $\Pi = 0$, $(dt/dr)_{\text{null}}$ [using the coordinate choice II(31)], with the slope of this boundary, $(dt/dr)_\Pi$, obtained from (A2b) (see Sec. IV and Fig. 8). From Eqs. (9) and (A2b), these slopes are given by

$$\left[\frac{dt}{dr}\right]_\Pi = \pm \frac{1}{L^{1/2}}\left[\frac{r^2}{3r^2 + L_0/L - 6\alpha_0}\right]^{1/2}, \tag{A3a}$$

$$\left[\frac{dt}{dr}\right]_{\text{null}} \approx \pm \frac{1}{L^{1/2}}, \tag{A3b}$$

so that $\Pi = 0$ is spacelike or null if $|(dt/dr)_{\text{null}}| \geqslant |(dt/dr)_\Pi|$ holds [i.e., condition (10)].

From Eqs. (A1c) and (A2a), $Q = 0$ is incompatible with $t = 0$, and so the expansion Θ cannot be zero along this boundary. However, from Eqs. (A2b) and (A3), the behavior of Θ and so the evolution of the fluid layers depends on the sign of the quantity

$$\beta_0 \equiv L_0/L - 6\alpha_0. \tag{A4}$$

If $\beta_0 > 0$, then the domain of regularity, as shown in Fig. 21a, does not include $t = 0$. Hence, the expansion Θ does not vanish, implying that the fluid always expands or contracts (no bounces). Comparing Eqs. (A3), condition (10) holds (with inequality sign) for all r, and so the boundary $\Pi = 0$ is everywhere spacelike (see Fig. 8). This boundary is then a future/past (depending on the sign of Θ) spacelike null infinity surface $\mathscr{I}_\pm$ associated with particle horizons and analogous to that of the de Sitter solution (see Fig. 11).

The $\mathbb{R}^4$ topology of the Wyman solution, together with the monotonously expanding or contracting behavior of the fluid, would suggest the existence of another null infinity surface $\mathscr{I}$ associated with points at an infinite spacelike separation, roughly corresponding to the limit $r \to \infty$. Such a $\mathscr{I}$ surface should be null, since from Fig. 22(a), all surfaces Σ_t for $|t| > L_0^{1/2}/3L$ extend between the spacelike FD singularity $Q = 0$ and the spacelike $\mathscr{I}$ marked by $\Pi = 0$, corresponding to the infinite future (or past) of comoving observers ($\tau \to \pm\infty$). Hence, this $\mathscr{I}$ surface must be a null hypersurface emerging as the limit of timelike and spacelike hypersurfaces Σ_r and Σ_t as $r \to \infty$ and $t \to \infty$ [i.e., $v \to \infty$ with v defined by Eq. I(37d)]. However, the (t,r) coordinate representation is not very helpful to study the behavior of $\mathscr{M}$ in this limit, and so, as illustrated by Fig. (22b), the (τ, r) coordinates introduced in Sec. X of Part II [Eqs. II(69)] are better suited for this purpose. For an expanding configuration ($\Theta > 0$) in these coordinates, the boundary $\Pi = 0$ in Fig. 22(a) is mapped towards $\tau \to \infty$ in Fig. 22(b), and so it is possible to use Eqs. II(70) and II(71) to explore the behavior of null geodesics as $r \to \infty$ along hypersurfaces of con-

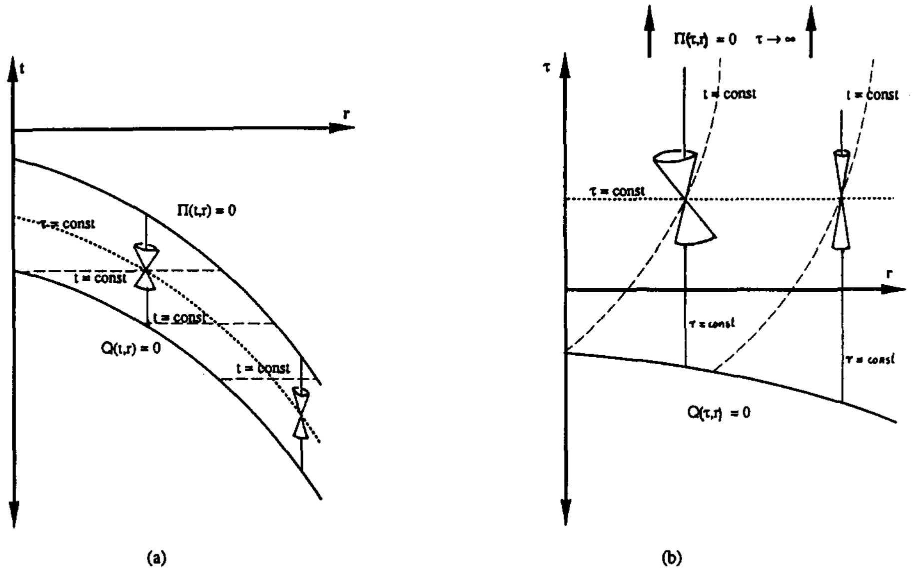


FIG. 22. Asymptotic behavior of null geodesics as $r \to \infty$. As illustrated in (a), the coordinates (t,r) are not very helpful to examine the behavior of null geodesics as $r \to \infty$, as surfaces Σ_t do not extend along infinite values of r but hit the boundary $\Pi = 0$. Since this boundary is at the infinite future of comoving observers, in the (τ,r) representation introduced in Eq. II(69a), it is mapped to infinite values of τ with surfaces Σ_t bending upwards, as shown in (b). The behavior of null geodesics at the limit $r \to \infty$ can be examined along surfaces of constant τ by using Eqs. II(70) and II(71) [see Fig. 5(c) of Part II]. This can be done for all particular cases of the Wyman solutions (and other solutions as well, see Fig. 19). Since Λ defined by Eq. II(69) diverges in this limit [the slopes of the light cones become "vertical" in (b)], for all cases of the Wyman solution, one can infer that there is a $\mathscr{I}$ surface as $r \to \infty$, as a null limit of timelike surfaces Σ_r. See Fig. 23.

stant τ [see Figs. 22(b)]. Since H and Λ [defined by Eq. II(69b)] diverge as $r \to \infty$, the slopes of the light cones become "vertical" in this limit, implying that the world lines of comoving observers do tend to a null hypersurface as $r \to \infty$. Since the other boundaries, such as the FD singularity and $\Pi = 0$ are everywhere spacelike. $\mathscr{M}$ is in this case globally hyperbolic, and from the information provided by Figs. 21(a) and 22, it is possible to infer qualitatively the form of its conformal Penrose diagram. This tentative diagram is displayed in Fig. 23(a).

If $\beta_0 \leqslant 0$, as shown by the domain of regularity in Fig. 21(b), the boundary $\Pi = 0$ hits $t = 0$ at $r_1 = |\beta_0/3|^{1/2}$. Since Θ vanishes at $t = 0$, fluid layers labeled by $0 \leqslant r < r_1$, emerging from the FD singularity in their past, bounce at $t = 0$ and fall back into this singularity in their future [see Fig. 21(b)]. Comoving observers with $r > r_1$ do not bounce and reach $\Pi = 0$ in their infinite future. However, the boundary $\Pi = 0$ is only spacelike [from Eqs. (A3)] for $r > r_2 = |\beta_0/2|^{1/2}$, becoming null at $r = r_2$, and timelike for $r_1 < r < r_2$, so that $\mathscr{M}$ is no longer globally hyperbolic. This situation is an example of restrictions in the domain of regularity due to a choice of equation of state (i.e., supplementary regularity conditions, see Sec. VIII of Part II). Collins did comment that the parameters of the Wyman solution allow for Θ to vanish, although he did not study this possibility in detail (see the end of his Sec. IV). As in the case discussed in the previous paragraph, the existence of a null $\mathscr{I}$ surface can be also inferred from the behavior of null geodesics as $r \to \infty$ along hypersurfaces of constant τ (see Fig. 22). A tentative conformal diagram for the type (ii) Wyman solution in which Θ vanishes is depicted in Fig. 23(b).

2. Type (iv)

In this case, the boundaries of $\mathscr{M}$ marked by finite coordinate values are the AD big bang $H = 0$ and $\Pi = 0$ if $L > 0$, or only $H = 0$ if $L = 0$ (see Table II of Part II). In either case, causal curves are complete at these boundaries (see Sec. X of Part II and Sec. IV), and so (apparently) $\mathscr{M}$ is a complete manifold. The domain of regularity for both cases is shown in Fig. 21. Considering first the case $L > 0$, the coordinate representation of the boundary $H = 0$ in (t,r) coordinates is the same as Eq. (A2a) [see Figs. 21(a) and 21(b)], while $\Pi = 0$ coincides with Eq. (A2b), except that α_0 now has a different value given by

$$\alpha_0 = \mathrm{cn}^{-1}(0) > 0. \tag{A5}$$

The AD big bang is everywhere timelike (see Fig. 3 of Part II), but the conformal nature of $\Pi = 0$ must be found again by comparing slopes through condition (10) and Fig. 8. The slopes of null cones near $\Pi = 0$ and the slope of this boundary in (T,r) coordinates are the same as in Eq. (A3) with α_0 given by Eq. (A5). The conditions on the parameters leading to the occurrence of a bounce as $\Theta(t) = 0$ for $t = 0$ also depend on the sign of β_0 [with α_0 given by Eq. (A5)]. Therefore, if $\beta_0 > 0$, Θ does not vanish (fluid layers monotonously collapse or expand without bouncing), the boundary $\Pi = 0$ is everywhere spacelike, and as in the type (ii) solution, it marks a spacelike $\mathscr{I}$ surface as depicted by Fig. 11 (see Sec. IV).

If $\beta_0 < 0$ with α_0 given by Eq. (A5), there is a bounce at $t = 0$, with the domain of regularity as in Fig. 21(b). Considering a fluid configuration initially expanding from the AD big bang $H = 0$ in the infinite past of the comoving observers, as shown in Fig. 21(b), only those fluid layers labeled by $r > r_1$ [α_0 given by (A4)] reach the boundary $\Pi = 0$ in their infinite future, while comoving observers with $0 \leqslant r \leqslant r_1$ bounce at $t = 0$, collapsing asymptotically in the AD big bang in their infinite future. As in the type (ii) solution, the boundary $\Pi = 0$ is timelike and null for $r_1 < r < r_2$ and $r = r_2$, respectively. Again, as with the case of type (ii) solutions, by looking at the behavior of null geodesics as $r \to \infty$ along hypersurfaces of constant τ (see Fig. 22), it can be inferred that spacelike and timelike hypersurfaces Σ_t and Σ_r (which always reach $\Pi = 0$ if the fluid does not bounce) tend to a null $\mathscr{I}$ surface in the limit $r \to \infty$ and $t \to \infty$. The qualitatively form of the conformal diagrams of the Wyman solution in

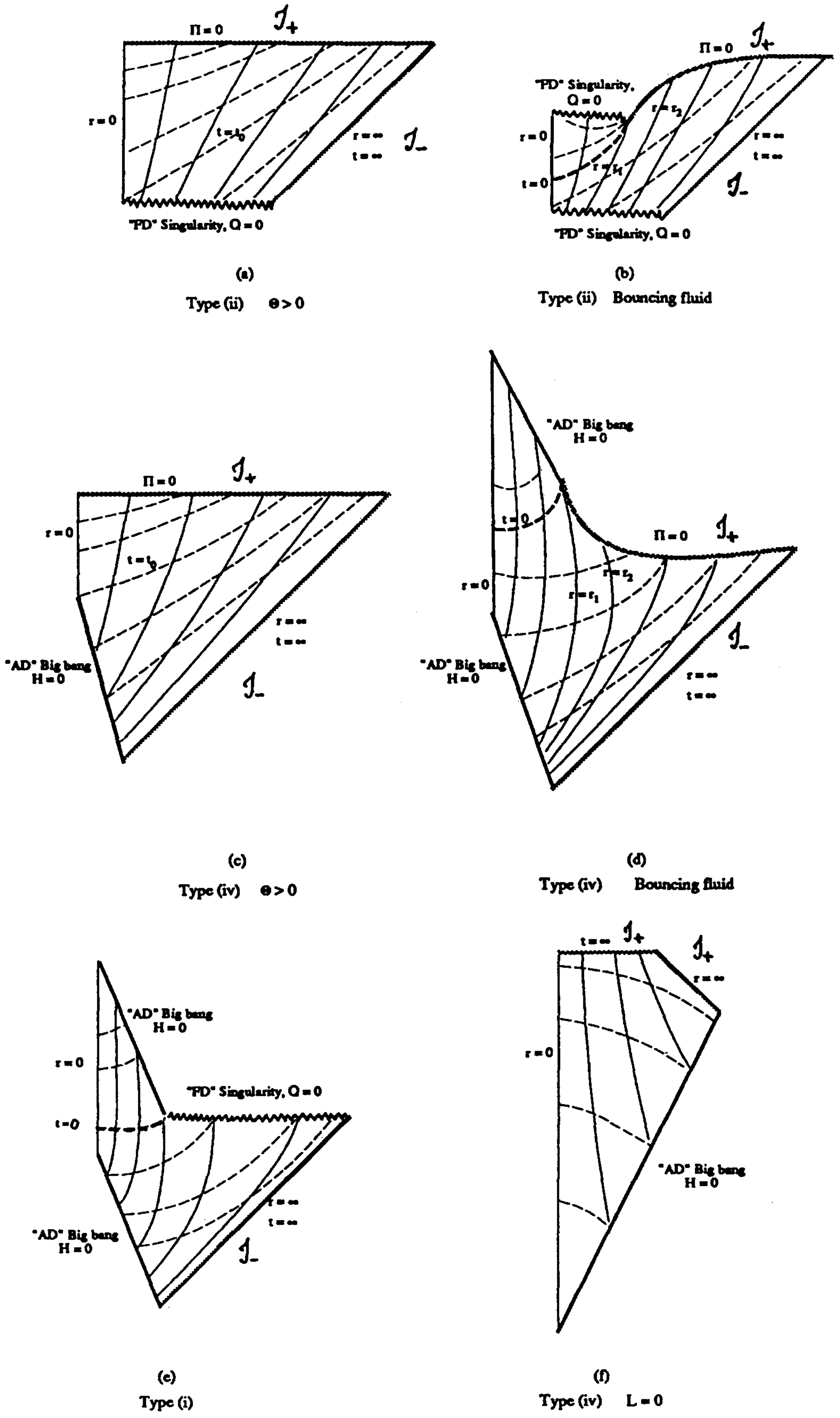


FIG. 23. Qualitative conformal diagrams for the various cases of the Wyman solution. Surfaces Σ_r and Σ_t are depicted as vertical and horizontal curves. Notice that the boundary $\Pi = 0$ is a $\mathscr{I}$ surface, which in the cases corresponding to (a) and (c) is everywhere spacelike. The existence of a second $\mathscr{I}$ surface has been inferred from the arguments described in Fig. 22.

the cases described above are shown in Figs. 23(c) and 23(d).

For the case in which $L=0$ [i.e., the NMcV($r3$)($X5$) solution, see Appendix A and Table II of Part II], $\Theta = \pm (L_0)^{1/2}$ is a nonzero constant, and so $T=t$. The AD big bang is marked by the same coordinate values as in Eq. (A1a) (with $L=0$), while there is no boundary $\Pi = 0$ labeled by finite coordinate values [see Fig. 21(c) for the domain of regularity]. However, information about the global structure of $\mathscr{M}$ can be obtained from a qualitative examination of the radial null geodesic equation II(56), which for this solution, takes the following particularly simple form:

$$\left[\frac{dt}{dr}\right]_{\text{null}} = \pm\left[\frac{L_0 H^3}{2|\mu|}\right]^{1/2} = \pm\left(\frac{L_0}{2}\right)^{1/2}|\mu|^{5/2}\left[t+\frac{r^2}{2}\right]^3. \tag{A6}$$

Consider a fluid configuration initially expanding from the AD big bang in the infinite past of the comoving observers. Since H and R diverge for finite r only if $t\to\infty$ corresponding to the infinite future of the fluid ($\tau\to\infty$) and the asymptotically de Sitter behavior $H\sim\exp\tau$, from Eq. (A6), the surface Σ_t corresponding to $t=\infty$ marks a spacelike $\mathscr{I}$ boundary in the infinite future analogous to $\Pi=0$ in the cases when $|\Theta|>0$ discussed above. Also, using the coordinate representation (τ,r), from Eqs. II(70) and II(71), if $r\to\infty$ along surfaces of constant τ, one has $(d\tau/dr)\to\pm\infty$, indicating that a null $\mathscr{I}$ boundary arises as a null limit of the surfaces Σ_r as $r\to\infty$. This information leads to the possible Penrose diagram shown in Fig. 23(f).

3. Type (I)

In this case, $L<0$ [see Table II(I)] and so the boundary $\Pi=0$ does not arise. The boundaries of $\mathscr{M}$ marked with finite coordinate values are then the AD big bang $H=0$ and the FD singularity $Q=0$. Since the coordinates of these boundaries are (A2a) and (A2b) with $\alpha_0 = \text{cn}^{-1}(0)>0$, respectively, the domain of regularity is as in Fig. 21(b). As β_0 is always negative, $t=0$ lies always within the domain of regularity of the solution, and so fluid layers with $0\leqslant r<r_1$ must necessarily bounce. Therefore, these layers either expand initially from the AD big bang in their infinite past and bounce back to this boundary in their infinite future (Θ passes from positive to negative), or contract initially from the FD singularity in their past bouncing back to it in the future (Θ passes from negative to positive). As in the cases discussed in previous paragraphs, there is also a null $\mathscr{I}$ boundary at the limit $r\to\infty$ and $|t|\to\infty$ (see Fig. 22). A tentative Penrose diagram for this case is depicted in Fig. 23(e).

APPENDIX B: THE McVITTIE SOLUTION

What is known in the literature as the McVittie solution (Refs. 9, 10, 21, and 44) corresponds to the neutral subclass of ChKQ solutions NMcV($r2$)($X2$), with the restriction $c=0$ [particular cases of solutions discussed in Sec. V, with functions h and X given by Eq. (11b)]. Using the choice of time coordinate II(33), this subclass corresponds to the metric II(33) with

$$H(T,r) = (1/b^2T)\,[T+(\mu/2)\bar h_{(2)}]^2, \tag{B1a}$$

$$(-g_{tt})^{1/2} = \frac{3}{\Theta}\left[b^2 - \frac{2\mu\bar h_{(2)}}{H}\right]^{1/2}$$

$$= \frac{\mathrm{T}-(\mu/2)\bar h_{(2)}}{[T\Theta(T)/3]\,[T+(\mu/2)\bar h_{(2)}]}, \tag{B1b}$$

where $\bar h_{(2)}$ is given by (11b). The solutions associated with the metric (B1) are characterized by the constant free parameters: μ, $\Delta = b^2$ and $k=0,\pm 1$ (see Table VI of Part I). The case $k=0$ is the UD solution NMcV($r2$)UD, while $\rho=\rho(T,r)$ if $k=\pm 1$. Following the current terminology in the literature, and in order to avoid confusion with the McVittie-type solutions introduced in Sec. IV of Part II, the above-mentioned subclass of ChKQ solutions will be referred in this Appendix as different particular cases of the McVittie solution.

In McVittie's pioneering paper,[9] as well as in subsequent literature,[21–23] it is suggested that McVittie's solution describes a "point particle" immersed in a perfect fluid which becomes asymptotically FRW. This characterization follows by writing the metric (B1) in "isotropic coordinates" (see Appendix C of Part I) as

$$ds^2 = \frac{[T-(\mu/b^{1/2}/x)(1+kx^2/4)^{1/2}]^2}{[T+(\mu/b^{1/2}/x)(1+kx^2/4)^{1/2}]^2}\,\frac{(dT/T)^2}{(\Theta(T)/3)^2}$$
$$+\frac{1}{b^4T^2}\left[T+\frac{\mu/b^{1/2}}{x}\left(1+\frac{kx^2}{4}\right)^{1/2}\right]^4$$
$$\times[dx^2+x^2(d\theta^2+\sin^2\theta\,d\phi^2)], \tag{B2}$$

so that identifying $\mu/b^{1/2}$ as the "mass" of the point particle and considering (T,x) values $x\approx 0$ and $\Theta(T)\approx$const, the term $\mu/b^{1/2}x$ dominates over the term $(1+kx^2/4)^{1/2}$, and the metric (B2) looks like a sort of nonstatic approximation to that of the Schwarzschild solution in "isotropic coordinates." Also, for large values of x the whole product $(\mu/b^{1/2}x)(1+kx^2/4)^{1/2}$ becomes negligible in comparison with T, and so, it is argued, (B2) approximates a FRW solution in the limit $x\to\infty$.

However, such coordinate-dependent identifications must be reexamined having in mind the global characteristics of the solutions discussed in Sec. V. For example, the locus $x=0$, which is supposed to be the world line of McVittie's point particle, corresponds to $r=0$ which, as shown in Sec. V, marks a null $\mathscr{I}$ boundary at infinite affine parameter distance along the surfaces Σ_T. This fact can be verified also from the radial null geodesic equation II(56), which for the metric (B1) takes the simple form

$$\left[\frac{dT}{dr}\right]_{\text{null}} = \frac{\pm\,\Theta(T)/3[(2by)^{1/2}T+\mu/2]^3}{2b^3y[(2by)^{1/2}T-\mu/2]}, \tag{B3}$$

so that $(dT/dr)_{\text{null}}\to\pm\infty$ as $r\to 0$ (i.e., as $y\to 0$ and/or $x\to 0$), indicating that $r=0$ is indeed a null boundary. However, even within the framework of the Schwarzschild solution in isotropic coordinates, the value $x = 0$ (with x being the "radial" isotropic coordinate of this solution) does not label the Schwarzschild singularity (i.e., point particle at $R = 0$) but points at infinity ($R \to \infty$) (see Ref. 57). Although McVittie's solution belongs to case (13c) (see Sec. V), implying that the null boundary at $r = 0$ is regular, with ρ and p taking the asymptotical values of equations (14), the characterization of this locus as the world line of a point particle is obviously inconsistent. Other global aspects of various particular cases of the McVittie solution will be discussed below. The cases $k = 0, \pm 1$ will be treated separately.

1. Case $k=1$

In this case, $f = \sin r$ and so there is a regular center at $r = \pi$, while, from Eq. (B3), $r = 0$ marks a regular null $\mathscr{I}$ boundary [ρ and p as $r \to 0$ are given by Eq. (14)]. As discussed in Sec. V, the space-time manifold $\mathscr{M}$ is in this case homeomorphic to $\mathbb{R}^4$. From Eqs. (B1), both $(-g_{tt})^{1/2}$ and H diverge as $T \to 0$ for r finite. Therefore $R = fH$ and $\tau = \int (-g_{tt})^{1/2}\, dT$ also diverge, with H, ρ, and p along comoving observers approaching $T = 0$ taking the "asymptotically de Sitter" behavior associated with $H \sim \exp[\Theta(0)\tau/3]$ and Eqs. (7b) and (7c) with $L = 0$. From Eq. (B3), the boundary $T = 0$ is spacelike if $|\Theta(0)| > 0$ and null if $\Theta(0) = 0$, and so it is analogous to the boundary $\Pi(T,r)$ discussed in Sec. IV, coinciding in this case with a single Σ_T surface (see Table II).

From Table VI of Part I, the solution is of type (ii) if $\mu > 0$, and of type (iv) if $\mu < 0$. In the first case, as illustrated by Fig. 24(a) the domain of regularity can take the following two forms:

$$T > (\mu/2)\bar{h}_{(2)}\,, \tag{B4a}$$

$$0 < T < (\mu/2)\bar{h}_{(2)}\,, \tag{B4b}$$

where, from Eqs. (10b) and I(16b), $\bar{h}_{(2)} = [2b^{1/2}\sin(r/2)]^{-1}$, and the FD singularity is marked by the coordinate values $T = (\mu/2)\bar{h}_{(2)}$. From Eq. (B3), and following Fig. 2 of Part II, this singularity is spacelike.

In the case (B4a), the fluid layers evolve from (toward) infinity ($R \to \infty$ for $0 < r < \pi$) toward (from) the FD singularity depending if $\Theta(T) < 0$ or $\Theta(T) > 0$. If $\Theta(T) = 0$, the fluid bounces, however, from Fig. 24(a), for whatever choice of $\Theta(T)$ (choice of equation of state) this singularity can never be avoided by all comoving observers. Also, the surfaces Σ_T reach the null $\mathscr{I}$ at $r = 0$ as $T \to \infty$, indicating that this null boundary is the null limit of timelike and spacelike surfaces Σ_r ($r \to 0$) and Σ_T ($T \to \infty$), as depicted by Fig. 12(b).

In the case (B4b), the evolution of the fluid, as illustrated by Fig. 24(a), is constrained between the asymptotically de Sitter boundary $T = 0$ and the FD singularity. Since, both boundaries are acausal (spacelike or null), $\mathscr{M}$ is globally hyperbolic. The basic difference between this case and that of (B4a) is that all surfaces Σ_T reach $r = 0$, and that the FD singularity can be avoided by all comoving observers if the fluid, initially contracting ($\Theta < 0$) from $T = 0$, bounces as $\Theta(T) = 0$ for $0 < T < \mu/4b^{1/2}$ [see Fig. 24(a)].

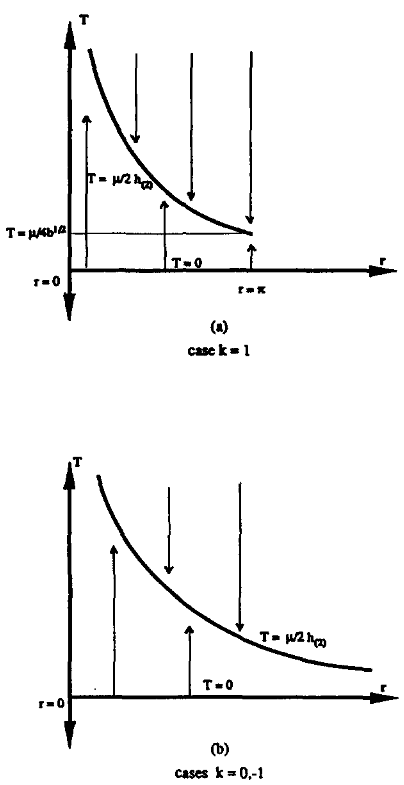


FIG. 24. Domain of regularity for the various cases of the McVittie solution. The vertical arrows indicate a monotonously collapsing ($\Theta < 0$) motion of fluid layers. If the latter were expanding ($\Theta > 0$), the arrows would point in the reverse direction.

If $\mu < 0$, the solution is of type (iv) and the domain of regularity is given by Fig. 24(a), except that now the coordinate constraint $T = \frac{1}{2}|\mu|\bar{h}_{(2)}$ marks the AD big bang at the infinite past/future of the comoving observers. Hence, if Θ is everywhere finite, $\mathscr{M}$ is a complete manifold but it is not globally hyperbolic since the AD big bang is a timelike boundary. If the domain of regularity is $T > \frac{1}{2}\mu\bar{h}_{(2)}$, no choice of Θ will make all fluid layers bounce and avoid $H = 0$. If Θ diverges along $T = T_0$, such that $H(T_0,\pi) = 0$, a null L singularity emerges, following a similar pattern as shown in Fig. 10. If Θ diverges for a value $T = T_1 \neq T_0$ [see Fig. 24(a)], one has a combination of L, FV singularities and AD big bang, as shown in Fig. 5 of Part II.

For the case (B4b), if $\Theta = 0$ for $0 < T < |\mu|/4b^{1/2}$, all fluid layers bounce and avoid the AD big bang. If Θ diverges for $T > |\mu|/4b^{1/2}$, one has a combination of L, FV singularities with AD big bang, as in Fig. 4 of Part II. Tentative conformal diagrams for some of the cases discussed above are shown in Fig. 25.

2. Cases $k=0$ and $k=-1$

If $k = 0,-1$, the McVittie solution has no regular center, and $\mathscr{M}$ is in all cases homeomorphic to $S^2 \times \mathbb{R}^2$, with the domain of regularity illustrated by Fig. 24(b) and given by Eqs. (B4) with $\bar{h}_2 = c_0/r$ if $k = 0$ [the UD solution NMcV($r2$)(UD) with $c_0 = b^{-1/2}$, see Table VII of Part I] or $\bar{h}_2 = [2b^{1/2}\sinh(r/2)]^{-1}$ if $k = -1$. As in the case

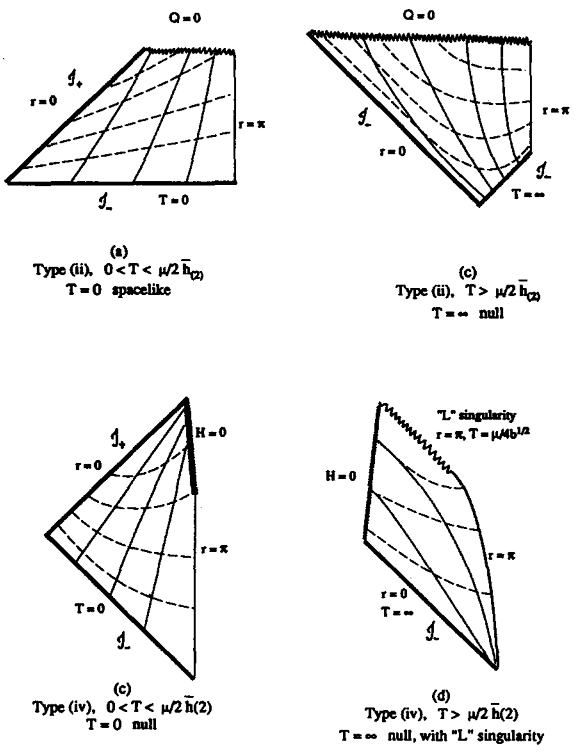


FIG. 25. Various possible conformal diagrams for the McVittie solution: Case $k=1$. As in Figs. 19 and 23, these conformal diagrams have been qualitatively sketched from the information available on the conformal structure of the boundaries and by looking at the behavior of timelike and spacelike surfaces Σ_r and Σ_T in the domain of regularity of Fig. 24. The latter three-surfaces are depicted as vertical and horizontal curves. It has been assumed that an equation of state is chosen so that $\Theta(T)$ is a monotonous function.

$k=1$, $r=0$ marks a regular null $\mathscr{I}$ boundary, while the boundary $T=0$ is associated to an asymptotically de Sitter behavior of H at the infinite future/past of comoving observers. From Eq. (B3), this boundary is spacelike in general, or null if $\Theta(0)=0$.

Although H and R diverge as $r\to\infty$ when the domain of regularity is (B4a) [see Fig. 24(b)], implying that points labeled by infinite values of r are at an infinite affine parameter distance along the Σ_T, the slopes of null cones behave as $(dT/dr)_{\rm null}\to\pm(\Theta/3)T^2$ as $r\to\infty$ [from Eq. (B3)]. Hence, unlike the case of the Wyman solution, the hypersurfaces Σ_r tend to a timelike boundary in the limit $r\to\infty$. Since Ω given by Eq. II(58) also diverges as $r\to\infty$, such a boundary is a timelike $\mathscr{I}$ surface, analogous to that of the universal covering of the anti-de Sitter solution. Thus, $\mathscr{M}$ is not globally hyperbolic in this case. On the other hand, if the domain of regularity is given by (B4b), then following the same arguments as in the Wyman solution, it can be inferred that a null $\mathscr{I}$ surface emerges at the limit $r\to\infty$ and $T\to 0$ [see Fig. (22)].

For $\mu>0$, the solution is of type (ii). If the domain of regularity is as in equation (B4a), the fluid layers terminate their evolution at the FD singularity $T=(\mu/2)\bar{h}_{(2)}$. The surfaces Σ_T only reach the boundary $r=0$ as $T\to\infty$, hence, this boundary appears as illustrated by Fig. 12(b), while there is a timelike $\mathscr{I}$ as $r\to\infty$. If the domain of regularity is as in Eq. (B4b), fluid layers evolve between the asymptotically de Sitter boundary $T=0$ and the FD singularity, with all surfaces Σ_T reaching the null boundary at $r=0$, and with a null $\mathscr{I}$ surface as $r\to\infty$ and $T\to 0$.

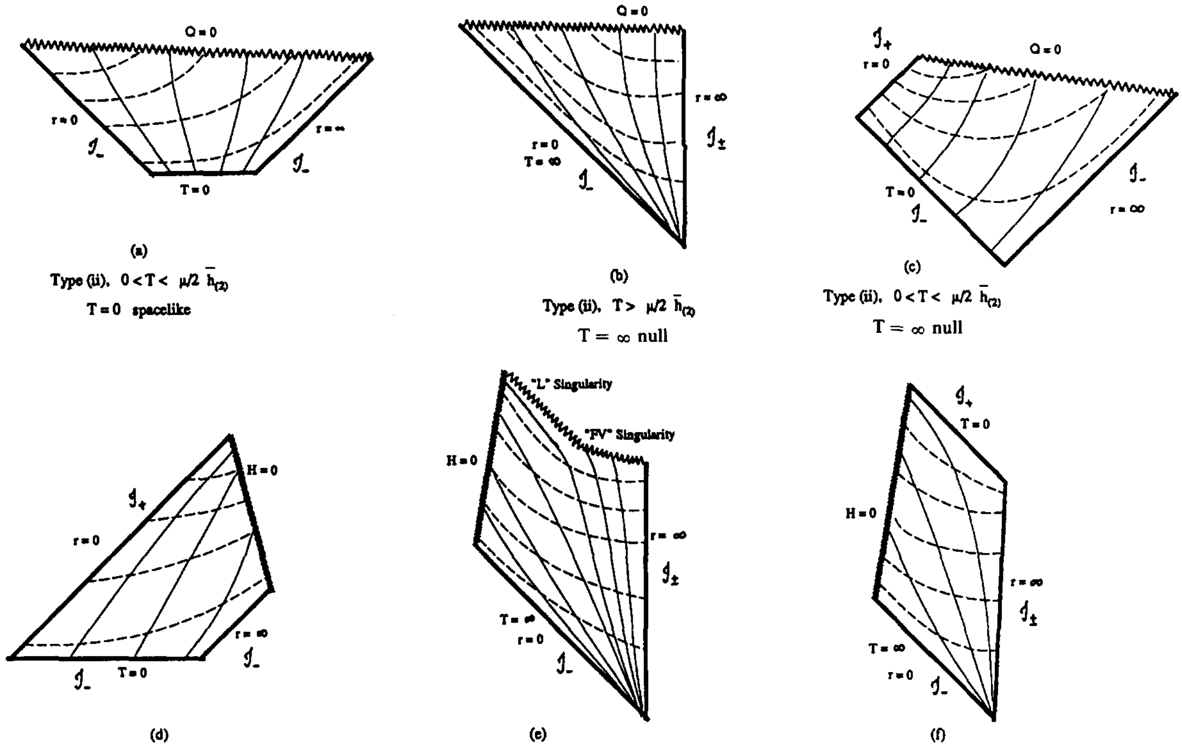
If $\mu<0$, the solution is of type (iv). As in the case of type (ii) solutions, the evolution of the fluid depends on the form of the domain of regularity. If Θ is everywhere finite, $\mathscr{M}$ is complete, while if Θ diverges for a given $T>0$, one has a combination of L and FV singularities and AD big bang, as illustrated by Fig. 5, Part II. The qualitative form of possible conformal diagrams for some of the cases discussed above are shown in Fig. 26.

3. Black holes and cosmological models

If the McVittie solution $k=1$ is matched to the Schwarzschild solution at $r=r_0$, so that the fluid region extends between r_0 and π, excluding $r=0$, one has a collapsing fluid sphere with a regular center, as described in Sec. XI of Part II. However, any other matching of this solution with a Schwarzschild space-time leads to a fluid region without a center, as discussed in Sec. VI. This fact was overlooked by McVittie[10] and Knutsen[11] in their study of the collapse of a NMcV($r2$)(UD) sphere (case $k=0$ of the McVittie solution), and by Mashhoon and Partovi[12] in a similar study of the charged version of this solution [the solution ChMcV($r2$)(UD), see Table VII of Part I]. Although, Knutsen's results concerning the formation of horizons and the evolution of the surface of the fluid sphere [i.e., basically a study of Eq. II(74a)] are correct, his reference to the singularity produced by the collapse of the surface of the sphere in the case $\mu<0$ [type (iv) solution] as a "central singularity" is misleading. First of all, the fluid region has no center, and as shown by Figs. 15(c), 15(d), and 16, the singularity produced in this case (the L singularity) is avoided by comoving observers labeled by $0<r<r_0$ (i.e., the interior layers), the latter evolving towards the AD big bang marked by $H=0$.

As mentioned earlier, the McVittie solution is inadequate to describe a point particle in a sort of asymptotically FRW background. As shown in this Appendix, the global features of McVittie's solution, for all cases $k=0$, ± 1, $\mu>0$, or $\mu<0$, bear no resemblance with what one would expect if it were a suitable characterization of such a model.[36] In fact, the cases $k=0,-1$ homeomorphic to $S^2\times\mathbb{R}$ have more in common, from a global point of view, to a Schwarzchild–Kruskal space-time than to a FRW cosmology. Hence, the resemblance of the metric (B2) with Schwarzchild's solution in isotropic coordinates could be connected with the fact that the static limit of the case $k=0$ is indeed the Schwarzschild solution (though, without the black-hole region $R<2m$, see Sec. X). Notice that for the cases corresponding to type (iv), for which $\mu<0$, the static limit is the Schwarzschild solution with negative mass.

The lack of a regular center in the cases with topology $S^2\times\mathbb{R}^2$, makes these solutions compatible with a Copernican principle which excludes privileged observers comoving along such symmetry centers. Though, the violation of energy conditions near the FD singularity, together with the fact



(a) Type (ii), $0<T<\mu/2\ \bar{h}_{(2)}$ $T=0$ spacelike

(b) Type (ii), $T>\mu/2\ \bar{h}_{(2)}$ $T=\infty$ null

(c) Type (ii), $0<T<\mu/2\ \bar{h}_{(2)}$ $T=\infty$ null

(d) Type (iv), $0<T<\mu/2\ \bar{h}_{(2)}$ $T=0$ spacelike

(e) Type (iv), $T>\mu/2\ \bar{h}_{(2)}$ $T=\infty$ null with "L" and "FV" singularities

(f) Type (iv), $T>\mu/2\ \bar{h}_{(2)}$ $T=\infty$ null

FIG. 26. Various possible conformal diagrams for McVittie solution: Cases $k=0,-1$. As in Fig. 25, these conformal diagrams are qualitative constructions. Surfaces Σ_r and Σ_T are depicted as vertical and horizontal curves. It has been assumed that an equation of state is chosen so that $\Theta(T)$ is a monotonous function.

that the AD big bang is timelike (just as the $\mathscr{I}$ boundary at $r=\infty$ in some cases) and so $\mathscr{M}$ is not globally hyperbolic, would probably exclude all cases from cosmological considerations. See Sec. XII.

APPENDIX C: EQUIVALENCE OF METRICS WITH DIFFERENT VALUES OF k IN SOLUTIONS WITH UNIFORM DENSITY

As shown in Appendix D of Part I, given the same combination of parameters (ϵ,μ,a,b,c,L), different values of k in Eq. II(2) denote different solutions if $\rho'\neq 0$. However, if $\rho'=0$, given the same combination of parameters (ϵ,μ,c_0,L), metrics with the radial coordinate given by the three different choices II(2) denote a single solution. This fact will be proved in this Appendix for the case $L=\text{const}$ using the choice of time coordinate II(33) (see Sec. V of Part II). The generalization of this result to $L=L(T)$ is straightforward.

If $\rho'=0$ and $dL/dT=0$, the metric II(1) can be expressed as

$$ds^2 = -c_0^4\left[1-\frac{2\mu c_0^5}{R}+\frac{\epsilon^2 c_0^6}{R^2}+Lc_0^2R^2\right]\frac{(dT/T)^2}{(\Theta/3)^2} + H^2\,dr^2 + R^2[d\theta^2+\sin^2\theta\,d\phi^2]\,, \tag{C1}$$

where $H=H(T,r)$ and $R(T,r)=fH, f(r)$ being either one of the functions given in Eqs. II(2). Keeping the same parameters (μ,ϵ,c_0, and L) and the same coordinate T, the metric (C1) for another choice of f is given by

$$ds^2 = -c_0^4\left[1-\frac{2\mu c_0^5}{\bar{R}}+\frac{\epsilon^2 c_0^6}{\bar{R}^2}+Lc_0^2\bar{R}^2\right]\frac{(dT/T)^2}{(\bar{\Theta}3)^2} + \bar{H}^2\,d\bar{r}^2 + \bar{R}^2[d\theta^2+\sin^2\theta\,d\phi^2]\,, \tag{C2}$$

where now $\bar{H}=\bar{H}(T,\bar{r})$ and $\bar{R}=\bar{f}(\bar{r})\bar{H}$, and so the barred quantities are functions of $\bar{r}$. For example, if $f(r)=r$ (choice $k=0$) then $\bar{f}(\bar{r})$ is either $\sin\bar{r}$ or $\sinh\bar{r}$ (choices $k=\pm 1$), and so on. If the metrics (C1) and (C2) are equivalent, the invariant quantities R and Θ must coincide; hence

$$\bar{\Theta}(T)=\Theta(T)\,, \tag{C3a}$$

$$\bar{f}(\bar{r})\bar{H}(T,\bar{r})=f(r)H(T,r) \tag{C3b}$$

must hold. If conditions (C3) imply

$$\bar{H}(T,\bar{r})d\bar{r}=H(T,r)dr\,, \tag{C4}$$

then (C1) and (C2) are equivalent with (C4) being the coordinate transformation $\bar{r}=\bar{r}(r)$ relating these metrics. From Eqs. I(38) and II(24), one has

$$\frac{f}{R}\frac{\partial R}{\partial r} = -c_0^2\,Q^{1/2} = \frac{\bar{f}}{\bar{R}}\frac{\partial\bar{R}}{\partial\bar{r}}\,, \tag{C5}$$

with Q given by (24b). Inserting (C3a) and (C3b) in (C5), Eq. (C4) follows immediately, proving the equivalence of (C1) and (C2). In order to generalize this result to solutions with $L = L(T)$, this function must be the same in both metrics, and the metric coefficient g_{tt} in Eqs. (14) and (24) must be written in terms of R.

[1]R. A. Sussman, J. Math. Phys. **28**, 1118 (1987); **29**, 945 (1988). These papers are referred to as "Part I" and "Part II," respectively.
[2]M. W. Cook, Aust. J. Phys. **28**, 413 (1975).
[3]A. Krasinski, Gen. Relativ. Gravit. **15**, 673 (1983).
[4]B. Mashhoon and H. Partovi, Ann. Phys. (NY) **130**, 99 (1980).
[5]B. Mashhoon and H. Partovi, Phys. Rev. D. **30**, 1839 (1984).
[6]C. B. Collins, J. Math. Phys. **29**, 2009 (1985).
[7]M. Wyman, Phys. Rev. **70** 396 (1946).
[8]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U. P., Cambridge, 1973).
[9]G. C. McVittie, Mon. Not. R. Astron. Soc. **93**, 325 (1933).
[10]G. C. McVittie, Astophys. J. **143**, 682 (1966); Ann. Inst. H. Poincaré **VI**, 1 (1967).
[11]H. Knutsen, Phys. Scr. **30**, 289 (1984).
[12]B. Mashhoon and H. Partovi, Phys. Rev. D **20**, 2455 (1979).
[13]E. Glass, J. Math. Phys. **20**, 1508 (1979).
[14]J. Eisenstaedt, Phys. Rev. D **11**, 2021 (1975); **12**, 1573 (1975), see also Ann. Inst. H. Poincaré **XXIV**, 179 (1976).
[15]H. A. Buchdahl, Astrophys. J. **140**, 1512 (1964).
[16]B. Mashhoon and E. Glass, Astrophys. J. **205**, 570 (1976).
[17]E. D. Fackerel, Astrophys. J. **153**, 643 (1968).
[18]J. R. Ray and J. C. Zimmerman, Nuovo Cimento B **42**, 183 (1977).
[19]J. R. Ray, Astrophys. J. **257**, 578 (1982).
[20]S. D. Maharaj, Ph.D. thesis, University of the Witwatwersrand, Johanesburg, South Africa, 1986.
[21]G. C. McVittie, *General Relativity and Cosmology* (Chapman and Hall, London, 1965), 2nd ed., pp. 181–186.
[22]P. D. Noerdlinger and V. Petrosian, Astrophys. J. **168**, 1 (1971).
[23]G. C. McVittie and R. P. A. C. Newman, Gen. Relativ. Gravit. **14**, 591 (1982).
[24]R. Geroch and G. T. Horowitz, in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U.P., Cambridge, 1979).
[25]R. Wald, *General Relativity* (University of Chicago Press, Chicago, 1984).
[26]B. O'Neil, *Semi Riemannian Geometry with Applications to General Relativity* (Academic, New York, 1983).
[27]J. Ehlers, in *Relativity, Astrophysics and Cosmology*, edited by W. Israel (Reidel, Dordrecht, Holland, 1973), pp. 114–116.
[28]C. J. S. Clarke, Class Quantum. Grav. **4**, L37 (1987).
[29]R. Penrose, in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U.P., Cambridge, 1979).
[30]R. Wald and P. Yip, J. Math Phys. **22**, 2659 (1981).
[31]W. C. Hernandez and C. W. Misner, Astrophys J. **143**, 452 (1966).
[32]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1972).
[33]C. Hellaby, Class Quantum. Grav. **4**, 635 (1987).
[34]P. Szekeres, Nuovo Cimento B **17**, 1 (1973).
[35]R. Gautreau, Nuovo Cimento **56**, 49 (1980).
[36]R. A. Sussman, Gen. Relativ. Gravit. **17**, 251 (1985).
[37]P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Scientists* (Springer, Berlin, 1971), 2nd. ed.
[38]J. B. Marion, *Classical Electromagnetic Radiation* (Academic, New York, 1965), Sec. 1.5.
[39]H. Stephani, Commun. Math. Phys. **4**, 137 (1967).
[40]I. H. Thomson and G. J. Whitrow, Mon. Not. R. Astron. Soc. **136**, 207 (1967); **139**, 499 (1968).
[41]H. Bondi, Mon. Not. R. Astron. Soc. **142**, 333 (1969).
[42]H. Knutsen, J. Math. Phys. **24**, 2188 (1983).
[43]W. B. Bonnor and M. C. Faulkes, Mon. Not. R. Astron. Soc. **137**, 239 (1967).
[44]D. Kramer, H. Stephani, M. A. H. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980).
[45]H. P. Künzle, Proc. R. Soc. London Ser. A **297**, 244 (1967).
[46]G. E. Tauber and J. W. Weinberg, Phys. Rev. **122**, 1342 (1962).
[47]J. Ehlers, P. Geren, and R. K. Sachs, J. Math. Phys. **9**, 1344 (1968).
[48]J. Ehlers, *General Relativity and Cosmology*, LXVII Enrico Fermi Summer School Proceedings, edited by R. K. Sachs (Academic, New York, 1971).
[49]G. F. R. Ellis, D. R. Matravers, and R. Treciokas, Gen. Relativ. Gravit. **15**, 981 (1983).
[50]W. A. Van Leewen, *On The Relativistic Kinetic Theory of Transport Phenomena* (Henkes Holland, N. V. Haarlem, 1971).
[51]R. Treciokas and G. F. R. Ellis, Commun. Math. Phys. **23**, 1 (1971).
[52]R. C. Tolman, Phys. Rev. **35**, 904 (1930).
[53]J. M. Stewart, *Non-equilibrium Relativistic Kinetic Theory* (Springer, Berlin, 1971).
[54]G. F. R. Ellis, *General Relativity and Cosmology*, LXVII Enrico Fermi Summer School Proceedings, edited by R. K. Sachs (Academic, New York, 1971).
[55]C. B. Collins, Can. J. Phys. **64**, 191 (1986).
[56]G. Darmois, Mem. Sci. Math. **XXV** (1927); G. Galloway, Ph.D. thesis, University of London, 1985; P. Vickers, Ph.D. thesis, University of London, 1973.
[57]R. Gautreau, Phys. Rev. D **29**, 198 (1984).

Spin-connection generalized Yang–Mills fields on double-dual generalized Einstein–Cartan backgrounds

G. M. O'Brien
School of Theoretical Physics, Dublin Institute for Advanced Studies, 10 Burlington Road, Dublin 4, Ireland

D. H. Tchrakian
Department of Mathematical Physics, St. Patrick's College, Maynooth, Kildare, Ireland and School of Theoretical Physics, Dublin Institute for Advanced Studies, 10 Burlington Road, Dublin 4, Ireland

(Received 11 August 1987; accepted for publication 30 December 1987)

In every even dimension, a modification of the generalized Yang–Mills systems (modGYM) is introduced. It is shown that spin-connection modGYM fields, which are self-dual on a double-self-dual background of generalized Einstein–Cartan (GEC) gravity, can be constructed—this generalizes the constructions of Charap and Duff [Phys. Lett. B **69**, 445 (1977)] in four dimensions, to all even dimensions. Additional duality constraints are also given that must be satisfied by the gravitational fields, if these constructions are to apply also to the full GYM systems. Applications to compactification, as well as relevant physical criteria, are briefly discussed.

I. INTRODUCTION

A systematic program to construct gauge field systems in higher dimensions, generalizing the Yang–Mills (YM) system, has been undertaken some time ago.[1,2]

Originally, the guiding criterion used was the presence of instantons in these systems. The $4p$-dimensional analog of the Belavin–Polyakov–Schwarz–Tyupkin (BPST) instantons[3] were found,[2,4] as well as the axially symmetric instantons[5] analogous to those found by Witten[6] for the YM system in R^4.

On the other hand, the physical criterion used was that under (coset space) compactification[7,8] these higher-dimensional systems yield residual systems in four dimensions whose low energy limits coincide with the YM–Higgs (YMH) system. Indeed, using[9] the calculus developed in Ref. 8, some such examples[10,11] have been considered in some detail, and we have referred to them as generalized YM (GYM) systems because they satisfy this criterion.

In connection with both the instanton solutions[2,5] in $4p$ dimensions and the properties of the vacuum structure in the (compactified) residual gauge field system,[10,11] a very central role is played by the *generalized* self-duality equations[1] pertaining to the GYM system in question. These duality equations solve the Euler–Lagrange equations of the GYM system. Some of these duality properties were considered also by Grossman *et al.*[12] and Brihaye *et al.*[13]

On a $2(p+q)$-dimensional manifold these duality conditions are

$$F(2p) = {}^*F(2p), \tag{1.1a}$$

$$^*F(2p)_{\mu_1\cdots\mu_{2p}} = (e/2q!)(i\kappa^2)^{q-p}\varepsilon_{\mu_1\cdots\mu_{2p}\nu_1\cdots\nu_{2q}}F(2q)^{\nu_1\cdots\nu_{2q}}, \tag{1.1b}$$

where e is the determinant of the *Vielbein* and $F(2p)$ is the totally antisymmetric p-fold product of the curvature two-forms $F(2)$. In general, for $p\neq q$, (1.1a) involves a nonzero power of the dimensional constant, and henceforth we will refer to such duality equations as *inhomogeneous*. By contrast, for $p=q$, the *homogeneous* duality equation (1.1a) involves no dimensional constant. As a result of the absence of κ in these cases with $q=p$, it was possible to find instanton solutions[2,5,13] of the GYM systems in R^{4p}. For $q\neq p$ there cannot exist any instantons on a flat background, but Bais and Batenburg[14] have made an extensive study for the case $p=1$ $(q>1)$ on $S^{2(p+q)}$, $CP^{(p+q)}$, and $HP^{(1/2)(p+q)}$. These field configurations[14] are very relevant, as examples, to the constructions given in the present paper.

The most obvious difference between the $p=q$ and $p\neq q$ cases is that (1.1a) holds in $4p$ dimensions in the first case, and in *every* even dimension in the second. The most important difference is that the GYM systems pertaining to (1.1a) are *conformally invariant* for the case $p=q$ but *not* for $p\neq q$, as is obvious, for example, due to the appearance of the parameter κ with dimensions of a length.

Another very interesting development has been the construction (in even dimensions) of the conformally invariant models by Zakrzewski[15] and others.[16] These systems do not concern us here directly, but are similar to the GYM in that they are endowed with self-duality conditions and, in $4p$ dimensions in particular, can be expressed as composite GYM field systems.

Having made our choice of gauge field systems, which we shall define in detail in Sec. III, the inevitable question that arises is, what is the gravitational partner of this gauge field? Clearly, in solving[14] self-duality equations such as (1.1), it is not necessary to specify the dynamics of the gravitational background. On the other hand, if we envisage that these gauge fields should model the fundamental interactions, it is desirable to specify their interactions with the gravitational background. This is the physical reasoning behind the considerations in the present paper, which in this sense can be considered complementary to the work of Bais and Batenburg.[14]

Before proceeding to consider gravity, we note the alternative extension of the YM system given by Saçlioğlu[17] and by Fujii,[18] and argue why we do not adopt them here.

As for GYM systems,[1,2] the definition of these[17,18] is

 0022-2488/88/051212-08$02.50

based on the existence of BPST configurations—not just in $4p$ but in every even dimension:

$$A_\mu = (2a/x(x^2+a^2))\Sigma_{\mu\nu}x_\nu, \tag{1.2a}$$

$$\Sigma_{\mu\nu}^{(\pm)} = -\tfrac{1}{4}((1\pm\Gamma_{d+1})/2)[\Gamma_\mu,\Gamma_\nu], \tag{1.2b}$$

where Γ_μ are the $2^{d/2}\times 2^{d/2}$ Γ matrices, Γ_{d+1} the chirality matrix, and $\Sigma_{\mu\nu}^{(\pm)}$ the spinor representation of so(d). This notation will be used throughout.

It turns out that the fields (1.2a) satisfy the equations of motion of the conformally invariant system

$$\operatorname{tr}(\Sigma_{\mu\nu}^{(\pm)}\otimes F_{\mu\nu})^{d/2}. \tag{1.3}$$

If all terms in (1.3) exhibiting higher than the second power of the velocity $\partial_\mu A_\nu$ are discarded,[17] then (1.3) coincides with GYM in $4p$ dimensions. On the other hand, in $(4k+2)$ dimensions (1.3) is not positive definite and under compactification does not reduce to a residual system that includes YMH at low energy. This is too high a price to pay for the conformal invariance of (1.3) in contrast to the absence of this symmetry in the case of GYM in $(4k+2)$ dimensions when $p\neq q$. Accordingly, we restrict our attention to GYM systems henceforth.

Perhaps the most popular scheme of coupling YM to gravity is through the Kaluza–Klein[15] (KK) mechanism. Here the gravitational system, namely the Einstein–Hilbert (EH) system, is taken to be the more fundamental. As such, in the higher dimensions the dynamics is given by the gravitational system alone, and after compactification[19,20] one finds the Einstein–YM system in four dimensions. Thus starting with the EH system whose Lagrange density exhibits *one* power of the Riemann curvature, one ends up with a residual system in which EH gravity interacts with YM, and the latter exhibits the *second* power of the gauge curvature. What then is the gravitational system which, when subjected to a KK compactification, yields a GYM system, say, in $4p$ dimensions? Clearly, the required system must involve the pth power of the Riemann curvature.

According to this reasoning, a direct generalization of EH gravity in higher dimensions was proposed[21,22] so that the Riemann curvature $R(2)$ in EH was replaced by a curvature $2p$-form $R(2p)$ in $4p$ dimensions. The $2p$-form $R(2p)$ is just the totally antisymmetric product of p-curvature two-forms, much in the same spirit in which $F(2p)$ in (1.1) was defined. We refer to the resulting gravitational systems as generalized Einstein–Cartan (GEC), because of the analogy with GYM. Indeed it was seen—for example, in Ref. 21—that under dimensional reduction these GEC systems give rise to residual systems that include the EH system, and that the latter dominate at low energy. These systems, which were first introduced by Lovelock,[23] have been arrived at also as string-generated gravitational systems,[24,25] and were applied in cosmological problems.[26]

On closer examination, it turns out that basing the analogy between GEC and GYM systems on a KK rationale[21,22] is false. The reason is that the KK *Ansatz* consists of imposing certain isometries on the *Vielbeine*, thus introducing Lie structure constants into the *Vielbein Ansatz*. These structure constants play the role of the generators of the Lie algebra in the residual system, where the YM system is just the square of the algebra-valued two-form $F(2)$. By contrast, the GYM system is the square of the $2p$-form $F(2p)$, which is certainly *not* algebra-valued. Therefore there is no reason to expect the KK *Ansatz*, which features only Lie structure constants, to generate a GYM system that cannot be expressed in algebra-valued quantities. To substantiate this assertion, we give a KK reduction of a $p=2$ GEC system in the Appendix, where we point out how the KK mechanism fails to generate a GYM system.

Abandoning a KK induced relationship between GEC and GYM means that both these geometric systems are to be treated on the same footing. Their relation, which will now amount to a duality between them, will be based on the occurrence of self-dual spin-connection GYM fields on double-self-dual GEC backgrounds. This generalizes the observation first made by Charap and Duff[27] for the EH–YM system in four dimensions, to the corresponding duality between GEC and GYM in every even dimension. In this process we shall find it useful to refine our original definition of GYM systems.[1,2]

This is the aim of the present work. In Sec. II we define the GEC systems and introduce the pertinent duality relations that solve the field equations. The corresponding definitions for the GYM systems and their stress tensors is presented in Sec. III. Putting these results together, we give the spin-connection fields in Sec. IV.

II. GENERALIZED GRAVITY (GEC)

We discuss the GEC gravitational systems before the GYM gauge systems because the former will turn out to be somewhat more fundamental than the latter, in spite of our abandonment of a KK mechanism where the gravitational system is clearly privileged. This situation will be clarified in Sec. IV.

We introduce our notation by defining the GEC system[23] on $d=2(p+q)$ dimensions, as given by the Lagrange density,

$$L(p,q) = \varepsilon^{\mu_1\cdots\mu_{2p}\nu_1\cdots\nu_{2q}}e^{a_1}_{\mu_1} \times e^{a_2}_{\mu_2}\cdots e^{a_{2p}}_{\mu_{2p}}\varepsilon_{a_1\cdots a_{2p}b_1\cdots b_{2q}}R^{\ b_1\cdots b_{2q}}_{\ \nu_1\cdots\nu_{2q}}, \tag{2.1a}$$

$$R(2q) = R^{\ b_1\cdots b_{2q}}_{\ \nu_1\cdots\nu_{2q}} = R^{\ [b_1b_2}_{\ [\nu_1\nu_2}R^{\ b_3b_4}_{\ \nu_3\nu_4}\cdots R^{\ b_{2q-1}b_{2q}]}_{\ \nu_{2q-1}\nu_{2q}]}, \tag{2.1b}$$

$$R(2) = R^{\ ab}_{\ \mu\nu} = \partial_{[\mu}\omega^{ab}_{\nu]} + \omega^{ac}_{[\mu}\omega^{cb}_{\nu]}. \tag{2.1c}$$

Here e^a_μ are the *Vielbeine*, whose determinants we have denoted as e in (1.1), and the greek and latin indices label coordinate and frame vectors, respectively. The $2q$-form $R(2q)$ is the q-fold totally antisymmetrized product of the curvature two-form $R(2)$, expressed as a curl of the spin connection ω^{ab}_μ in (2.1c). Clearly, when $p=0$ and $d=4q$, (2.1a) is a total divergence and hence dynamically trivial. For $p=q=1$, (2.1) is the EH system.

A. Equations of motion

The Euler–Lagrange equations resulting from the variations of e^a_μ and ω^{ab}_μ, respectively, are

$$t^{\mu}_{a}=2p\varepsilon^{\mu\mu_2\mu_3\cdots\mu_{2p}\nu_1\cdots\nu_{2q}}e^{a_2}_{\mu_2}\cdots e^{a_{2p}}_{\mu_{2p}}$$
$$\times\varepsilon_{aa_2a_3\cdots a_{2p}b_1\cdots b_{2q}}R^{\ b_1\cdots b_{2q}}_{\ \nu_1\cdots\nu_{2q}}=0, \tag{2.2a}$$
$$t^{\mu}_{ab}=q\varepsilon^{\mu\mu_2\cdots\mu_{2p}\nu_1\cdots\nu_{2q}}e^{a_2}_{\mu_2}\cdots e^{a_{2p}}_{\mu_{2p}}$$
$$\times\varepsilon_{aba_1\cdots a_{2p}b_3\cdots b_{2q}}T^{a_1}_{\ \nu_1\nu_2}R^{\ b_3\cdots b_{2q}}_{\ \nu_3\cdots\nu_{2q}}. \tag{2.2b}$$

Equation (2.2a) is the Einstein equation, and (2.2b) is the torsion equation exhibiting the torsion field

$$T^{a}_{\ \mu\nu}=D_{[\mu}e^{a}_{\nu]}, \tag{2.3a}$$
$$D_{\mu}e^{a}_{\nu}=\partial_{\mu}e^{a}_{\nu}+\omega^{ab}_{\mu}e^{b}_{\nu}. \tag{2.3b}$$

In terms of the pseudotensor t^{μ}_{a} in (2.2a), we can define an Einstein tensor of the curvature $2q$-form $R(2q)$ by

$$t^{\mu}_{a}=-2p2p!2q!eG^{\mu}_{a}(2q), \tag{2.4a}$$
$$G^{\mu}_{a}(2q)=R^{\mu}_{a}(2q)-(1/2q)e^{\mu}_{a}\mathscr{R}(2q), \tag{2.4b}$$
$$R^{\mu}_{a}(2q)=R^{\ \mu\nu_2\cdots\nu_{2q}}_{\ ab_2\cdots b_{2q}}e^{b_2}_{\nu_2}\cdots e^{b_{2q}}_{\nu_{2q}}, \tag{2.4c}$$
$$\mathscr{R}(2q)=R^{\mu}_{a}(2q)e^{a}_{\mu}. \tag{2.4d}$$

In the presence of another field interacting with gravity, (2.2a) will be replaced by the Einstein equation

$$G^{\mu}_{a}(2q)=\Theta^{\mu}_{a}(2q), \tag{2.5}$$

where $\Theta^{\mu}_{a}(2q)=e_{\nu a}\Theta^{\mu\nu}(2q)$ is the corresponding stress tensor.

Concerning the torsion equation (2.2b), there is a very important difference between the EH case $p=q=1$, and all other cases with $q>1$. In the EH case (2.2b) implies vanishing torsion, while for $q>1$ this is *not* so. This fact has been exploited[21] in a compactification scheme recently.

Perhaps the most important feature of the Einstein equation (2.2a) is that it leads to the vanishing of the pseudoscalar quantity $e\mathscr{R}(2q)$ given by (2.1a). The expression $\mathscr{R}(2q)$ is a function of the Ricci scalar $R(2)$, as well as the other components of the Riemann tensor $R(2)$. Consequently, in a compactification scheme characterized by the manifold $M_d=M_4xK^{d-4}$, where K^{d-4} is a compact space of constant curvature, the vanishing of $\mathscr{R}(2q)$ does not imply the constancy of $\mathscr{R}(2)$, the Ricci scalar on M_4. This feature for the example with $p=q=2$ was exploited in Ref. 21 for the case of nonvanishing torsion, but this last property is not necessary for the former feature to persist. We shall return to this point again below.

In the present work, we shall consider only those field configurations for which (2.2b) is satisfied with $T^{a}_{\ \mu\nu}=0$. We set the torsion equal to zero in anticipation of our use of double-duality conditions. We shall see below that double-duality conditions are not useful for solving (2.2) in the presence of torsion.

As a consequence of setting the torsion equal to zero, the covariant divergence of (2.2a) vanishes:

$$(1/e)D_{\mu}(eG^{\mu}_{a}(2q))=0, \tag{2.6a}$$
$$(1/e)D_{\mu}(ee^{\mu}_{a})=0, \tag{2.6b}$$

where we have recorded another useful consequence of vanishing torsion, in (2.6b).

We now turn our attention to duality conditions which solve the Einstein equation (2.2a).

B. (Single-)self-duality

Because in the study of the EH system in four (Euclidean) dimensions (single-)self-duality[28] of the Riemann curvature plays an important role, we consider it also for the GEC gravity. Algebraically it plays the same role; for example, in $4p$ dimensions we have

$$R^{\ a_1\cdots a_{2p}}_{\ \mu_1\cdots\mu_{2p}}=(e/2p!)\varepsilon_{\mu_1\cdots\mu_{2p}\nu_1\cdots\nu_{2p}}R^{\ \nu_1\cdots\nu_{2p}}_{\ a_1\cdots a_{2p}} \tag{2.7}$$

as the self-duality condition. (Contravariant and covariant frame indices are not distinguished in this Euclidean space.)

Substitution of (2.7) into (2.2a) yields

$$\varepsilon^{\mu\mu_2\cdots\mu_{4p}}D_{\mu_2}T_{\mu_3\mu_4,\mu_5}\cdots D_{\mu_{4p-2}}T^{a}_{\ \mu_{4p-1}\mu_{4p}}=0, \tag{2.8}$$

where $T_{\mu\nu,\rho}=T^{a}_{\ \mu\nu}e^{a}_{\rho}$. So for vanishing torsion, (2.8) is identically satisfied, and hence the self-duality condition (2.7) solves the Einstein equation (2.2a).

In practice, however, (2.7) seems to be too strong a constraint on the spin connection, and no such solutions are known except in the EH case[28] when $p=q=1$. The reason for this may be that the EH case ($p=1$) is distinguished over all others ($p>1$) in that for $p=1$ the self-duality of the spin connection

$$\omega^{ab}_{\mu}=(1/2!)\varepsilon^{abcd}\omega^{cd}_{\mu} \tag{2.9}$$

implies (2.7), and it is Eq. (2.9) that is actually solved.[28] For $p>1$, we have not succeeded in finding any analog of (2.9) that implies (2.7). For this reason, we restrict ourselves to the (weaker) double-self-duality conditions in the rest of this paper.

C. Double-self-duality

(1) $p=q$: In four dimensions, double-duality of the Riemann curvature

$$R^{\ mn}_{\ \mu\nu}=(e/2!^2)\varepsilon_{\mu\nu\rho\sigma}R^{\ \rho\sigma}_{\ rs}\varepsilon^{rsmn} \tag{2.10}$$

implies, in the absence of torsion, a field configuration satisfying the Einstein equations with a cosmological constant. This is easily seen by substituting (2.10) into the Einstein equation (2.2a), which in the notation of (2.4) results in the constraint

$$G^{\mu}_{a}(2)=-\tfrac{1}{4}e^{\mu}_{a}\mathscr{R}(2). \tag{2.11}$$

By virtue of Eqs. (2.6), the covariant divergence of (2.11) then implies that the Ricci scalar $\mathscr{R}(2)$ is a (cosmological) constant, as is well-known.

When $p=q>1$, that is, in $4p$ dimensions only, the double-duality condition on the Riemann curvature reads

$$R^{\ m_1\cdots m_{2p}}_{\ \mu_1\cdots\mu_{2p}}=(e/2p!^2)\varepsilon_{\mu_1\cdots\mu_{2p}\nu_1\cdots\nu_{2p}}$$
$$\times R^{\ \nu_1\cdots\nu_{2p}}_{\ n_1\cdots n_{2p}}\varepsilon^{n_1\cdots n_{2p}m_1\cdots m_{2p}}. \tag{2.12}$$

Substituting (2.12) into (2.2a) and using the notation of (2.4) similarly yields the constraint

$$G^{\mu}_{a}(2p)=-(1/4p)e^{\mu}_{a}\mathscr{R}(2p). \tag{2.13}$$

The covariant divergence of (2.13), again by virtue of Eqs. (2.6), implies the constancy of $\mathscr{R}(2p)=\eta$, say. This field configuration can be interpreted as the solution to the field equations of the system

$$L=L(p,p)-\tfrac{1}{2}2p!^2e\eta, \tag{2.14}$$

where $L(p,p)$ is given by (2.1a) with $q=p$, and is a generalized cosmological constant. The HP^2 solution of Ref. 22 is a $p=q=2$ example.

We note here that η is not equal to $\mathscr{R}(2)$, the Ricci scalar as in EH gravity, but rather to $\mathscr{R}(2p)$. The latter is a function of $\mathscr{R}(2)$, as well as the other components of the Riemann curvature $R(2)$. In the simplest case, for $p=2$, this is

$$\mathscr{R}(4)=\mathscr{R}(2)^2-4R_{\mu}^{\ a}R_{\ a}^{\mu}+R_{\mu\nu}^{ab}R_{ab}^{\mu\nu}=\eta. \quad (2.15)$$

Requiring in (2.15) the constancy of $\mathscr{R}(4)$ is different from requiring the constancy of $\mathscr{R}(2)$ ($=\mathscr{R}$) in (2.15). Although the condition (2.15) is fulfilled with a nonzero constant Ricci scalar for double-dual solutions known to us,[22] it is not inconceivable that there may be other solutions for which a vanishing Ricci scalar is consistent with (2.15) and its higher-order analogs (in higher dimensions). Such solutions, *if they exist*, could have very interesting applications to compactification. The difference from the previous scheme suggested in Ref. 21 is that here we are dealing with a *torsionless* connection.

(2) $q>p$: In this case we can write down, in *every even dimension*, the double-duality relations

$$R_{\mu_1\cdots\mu_{2p}}^{m_1\cdots m_{2p}}=(\kappa^2)^{q-p}(e/2q!^2)\varepsilon_{\mu_1\cdots\mu_{2p}\nu_1\cdots\nu_{2q}}\times R_{n_1\cdots n_{2q}}^{\nu_1\cdots\nu_{2q}}\varepsilon^{n_1\cdots n_{2q}m_1\cdots m_{2p}}, \quad (2.16)$$

and their inverses. Clearly, κ here is a constant with the dimensions of a length.

The dynamical system pertaining to this nonlinear duality relation is the following combination of GEC systems:

$$L_{\rm GEC}=\tau L(p,q)+(1/\lambda)L(q,p), \quad (2.17)$$

where τ, the constant coefficient, has the same dimensions as λ, and the latter has dimensions of $\frac{1}{2}(q-p)$th power of a length.

In the absence of torsion, we have only one equation of motion, (2.2a), the Einstein equation, which we could express in terms of the tensor $E_a^{\mu}=e^{-1}t_a^{\mu}$ as $E_a^{\mu}=0$. We reexpress this equation through

$$E_m^{\mu}=-2p!2q![2q\tau G_m^{\mu}(2q)+2p\lambda^{-1}G_m^{\mu}(2p)], \quad (2.18)$$

where the generalized Einstein tensors $G_m^{\mu}(2r)$ are defined by (2.4b)–(2.4d).

For the double-self-dual field configuration satisfying (2.16), E_m^{μ} of (2.18) takes the following form:

$$\begin{aligned}E_m^{\mu}=&[2q2p!\lambda^{-1}(\kappa^2)^{q-p}G_m^{\mu}(2q)\\&+2p2q!^2\tau(\kappa^2)^{p-q}G_m^{\mu}(2p)]\\&+e_m^{\mu}[2p!^2\lambda^{-1}(\kappa^2)^{q-p}\mathscr{R}(2q)\\&+2q!^2\tau(\kappa^2)^{p-q}\mathscr{R}(2p)].\end{aligned} \quad (2.19)$$

The consistency of (2.18) and (2.19) requires that the expression in the square brackets in (2.18) be proportional to that in the first square brackets in (2.19). This can be achieved by identifying the coefficients of $G_m^{\mu}(2q)$ and $G_m^{\mu}(2p)$, respectively. The resulting identities can only be satisfied if

$$2p!(\kappa^2)^{q-p}=2q!\tau\lambda. \quad (2.20)$$

In this case, (2.19) reads

$$E_m^{\mu}=2q!2p!\{[2q\tau G_m^{\mu}(2q)+(2p/\lambda)G_m^{\mu}(2p)]+e_m^{\mu}[\tau\mathscr{R}(2q)+(1/\lambda)\mathscr{R}(2p)]\}. \quad (2.21)$$

Equating (2.18) and (2.21) then yields

$$[2q\tau G_m^{\mu}(2q)+(2p/\lambda)G_m^{\mu}(2p)]+\tfrac{1}{2}e_m^{\mu}[\tau\mathscr{R}(2q)+(1/\lambda)\mathscr{R}(2p)]=0, \quad (2.22)$$

which is the equation analogous to (2.13) in the $p=q$ case. Indeed, for $p=q$, (2.13) and (2.22) coincide as expected, provided due note is taken of (2.20).

Again, because of Eqs. (2.6) we conclude from (2.22) that

$$\eta=\tau\mathscr{R}(2q)+(1/\lambda)\mathscr{R}(2p) \quad (2.23)$$

is a (cosmological) constant, and that therefore the double-self-dual field configuration (2.16) satisfies (2.22), the Euler–Lagrange equations of the following generalized Einstein system (with cosmological constant):

$$L=[\tau L(p,q)+(1/\lambda)L(q,p)]-\tfrac{1}{2}2p!2q!e\eta. \quad (2.24)$$

This also coincides with (2.14) when $p=q$. The only qualitative difference between the $p=q$ and $p<q$ cases is the constraints obtained by contracting the Einstein equations (2.13) and (2.22) with e_{μ}^{m}, respectively. In the former case this constraint is trivial, while in the latter it reads

$$\mathscr{R}(2p)=\tau\lambda\mathscr{R}(2q). \quad (2.25)$$

As expected, (2.25) is precisely the constraint that follows from the total contraction of the double-self-duality relation (2.16). It follows from (2.25) and (2.23) that

$$\mathscr{R}(2p)=\eta\lambda, \quad (2.26a)$$

$$\mathscr{R}(2q)=\eta/\tau. \quad (2.26b)$$

When $p=1$ and/or $q=1$, Eqs. (2.26) imply that the Ricci scalar $\mathscr{R}(2)$ is constant. Otherwise, the qualitative remarks we made after (2.14) apply.

III. GENERALIZED YANG–MILLS (GYM)

In this section we will repeat our definition[1,2] of the GYM systems. We will (a) *modify* these definitions in a manner that suits our purposes in Sec. IV, namely, that of generalizing the Charap–Duff[27] construction to all even dimensions. In addition, we will (b) show that the stress tensor of these GYM systems vanishes when the generalized self-duality relations (1.1a) hold.

In the construction of the GYM systems in $2(q+p)$ dimensions, the cornerstone is the duality relations (1.1a),

$$F(2p)={}^*F(2p), \quad (3.1a)$$

$$F(2q)={}^*F(2q), \quad (3.1b)$$

where ${}^*F(2p)$ is the Hodge dual of $F(2q)$ given by (1.1b). The expression (3.1b) is simply the inverse relationship to (3.1a), to which it is equivalent. That (3.1) minimizes the action

$$L_{\rm GYM}(p,q)=\int dx\, e\,{\rm tr}\left[F(2p)^2+(i\kappa^2)^{2(q-p)}\frac{2p!}{2q!}F(2q)^2\right] \quad (3.2)$$

is self-evident. The equations of motion are solved[1] explicitly by (3.1) and the Bianchi identities, and the action is then equal to the $(p+q)$th Chern–Pontryagin integral.

This action is conformally invariant only when $p=q$, and hence instanton solutions[2,5] on R^{4p} occur only for this case. In all other cases $q\neq p$, instanton solutions exist only on curved backgrounds, e.g., on $S^{2(p+q)}$, $CP^{(p+q)}$, and $HP^{(p+q)/2}$ given in Ref. 17.

It turns out that the generalization of the Charap–Duff mechanism, which is our most important aim here, is facilitated by considering a modification of the GYM systems. These are given immediately below.

A. The modified GYM

In (3.2) we have defined the $2p$-form $F(2p)$ as the totally antisymmetrized product of p factors of the curvature two-form $F(2)$. Instanton solutions[2,5] of this system occur in R^{4p}, when $q=p$ and $F(2)$ takes its values in the spinor representations of SO$(4p)$.

Our modification of the systems (3.2) pertains only to those cases when $F(2)$ takes its values in the algebra of SO$(2(p+q))$, and in particular in the $2^{p+q-1}\times 2^{p+q-1}$ spinor representations. The modification consists of the replacement of $F(2r)$ in (3.2) by $\hat{F}(2r)$:

$$\hat{F}(2r)=\hat{F}_{\mu_1\cdots\mu_{2r}}$$

$$=\{[(2r-1)(2r-3)\cdots 3]^2 r!\}^{-1}F^{m_1\cdots m_{2r}}_{\mu_1\cdots\mu_{2r}}\Sigma_{m_1\cdots m_{2r}}, \tag{3.3a}$$

$$F^{m_1\cdots m_{2r}}_{\mu_1\cdots\mu_{2r}}=F^{[m_1m_2}_{[\mu_1\mu_2}F^{m_3m_4}_{\mu_3\mu_4}\cdots F^{m_{2r-1}m_{2r}]}_{\mu_{2r-1}\mu_{2r}]}, \tag{3.3b}$$

$$\Sigma^{(\pm)}_{m_1\cdots m_{2r}}=\Sigma^{(\pm)}_{[m_1m_2}\Sigma^{(\pm)}_{m_3m_4}\cdots\Sigma^{(\pm)}_{m_{2r-1}m_{2r}]}. \tag{3.3c}$$

In (3.3b) and (3.3c) the square brackets on the indices signify total antisymmetrization, and $\Sigma^{(\pm)}_{mn}$ in (3.3c) refers to the elements of the algebra of SO$(2(p+q))$ in the spinor representations

$$\Sigma^{(\pm)}_{mn}=-\tfrac{1}{4}((1\pm\Gamma_{d+1})/2)[\Gamma_m,\Gamma_n], \tag{3.4}$$

where the Γ_m $(m=1,\ldots,2(p+q))$ are the $2^{p+q}\times 2^{p+q}$ Γ matrices and Γ_{d+1} is the chirality matrix in those dimensions. Finally, the $F^{mn}_{\mu\nu}$ in (3.3b) relate to $F(2)=F_{\mu\nu}$ through $F_{\mu\nu}=F^{mn}_{\mu\nu}\Sigma^{(\pm)}_{mn}$ as usual.

We characterize these modified GYM (modGYM) systems by the action $\hat{L}_{\rm GYM}(p,q)$ obtained from (3.2) by replacing $F(2p)$ and $F(2q)$ by $\hat{F}(2p)$ and $\hat{F}(2q)$ as given by (3.3):

$$\hat{L}_{\rm GYM}(p,q)$$

$$=\int dx\, e\, \mathrm{tr}\Big[\hat{F}(2p)^2+(i\kappa^2)^{2(q-p)}\frac{2p!}{2q!}\hat{F}(2q)^2\Big]. \tag{3.5}$$

In spite of the close similarity between (3.2) and (3.5), the dynamics of the two systems are quite distinct. Indeed, (3.5) is not minimized by the duality constraint (3.1), but by

$$\hat{F}(2p)={}^*\hat{F}(2p), \tag{3.6a}$$

$$\hat{F}(2q)={}^*\hat{F}(2q). \tag{3.6b}$$

The relationship between the GYM systems (3.2) and the modGYM systems (3.5) can be made transparent by finding the relation between $F(2r)$ and $\hat{F}(2r)$. We can express $F(2r)$ by

$$F_{\mu_1\cdots\mu_{2r}}=\{F_{\mu_1\mu_2},F_{\mu_3\mu_4},\ldots,F_{\mu_{2r-1}\mu_{2r}}\}$$

$$+(r-1)\text{ cycl. perms. on }(\mu). \tag{3.7}$$

The $(r-1)$ cyclic permutations range over the indices $[\mu_2\cdots\mu_{2r}],[\mu_4\cdots\mu_{2r}],\ldots,$ and $[\mu_{2r-2}\cdots\mu_{2r}]$, and the bracket[29] $\{A_1,A_2,\ldots,A_r\}$ denotes the sum of the $r!$ different orderings of $A_1,A_2,\ldots,A_r$.

With this same bracket notation,[29] we can express $\Sigma_{\mu_1\cdots\mu_{2r}}$ in (3.3c) as

$$\Sigma_{m_1\cdots m_{2r}}=\{\Sigma_{m_1m_2},\Sigma_{m_3m_4},\ldots,\Sigma_{m_{2r-1}m_{2r}}\}$$

$$+(r-1)\text{ cycl. perms. on }(m), \tag{3.8a}$$

which can be rewritten as

$$\{\Sigma_{m_1m_2},\Sigma_{m_3m_4},\ldots,\Sigma_{m_{2r-1}m_{2r}}\}$$

$$=\{[(2r-1)(2r-3)\cdots 3]\}^{-1}\Sigma_{m_1\cdots m_{2r}}+\mathbf{X}_{m_1\cdots m_{2r}}, \tag{3.8b}$$

where $\mathbf{X}(r)=\mathbf{X}_{m_1\cdots m_{2r}}$ is a tensor spinor that has increasingly complicated structure with increasing r.

Remembering definition (3.3) and that $F_{\mu\nu}=F^{mn}_{\mu\nu}\Sigma_{mn}$, we can put together (3.7) and (3.8b) to find the required relation

$$F_{\mu_1\cdots\mu_{2r}}=\hat{F}_{\mu_1\cdots\mu_{2r}}+\mathbf{X}_{m_1\cdots m_{2r}}(F^{m_1m_2}_{\mu_1\mu_2}\cdots F^{m_{2r-1}m_{2r}}_{\mu_{2r-1}\mu_{2r}}$$

$$+(r-1)\text{ cycl. perms. on }(\mu)). \tag{3.9}$$

From its definition in Eqs. (3.8), the computation of $\mathbf{X}(r)$ in terms of the invariant tensors $\delta^{n_1}_{m_1}$, $\delta^{n_1n_2}_{m_1m_2}$, etc., the unit matrix, and the spin matrices $\Sigma_{m_1n_1}$, $\Sigma_{m_1m_2n_1n_2}$, etc., is completely straightforward for any given r. Unfortunately, an expression for general r is not conveniently written, so it suffices here to give the $r=1,2,3$ cases explicitly. For $r=1$, which leads to the definition of the YM system, $\mathbf{X}_{mn}(1)=0$ and (3.2) coincides with (3.5). For $r=2,3$, on the other hand, we have

$$\mathbf{X}_{m_1m_2n_1n_2}(2)=-\tfrac{1}{3}\delta^{n_1n_2}_{m_1m_2}\mathbf{1}_8, \tag{3.10a}$$

$$\mathbf{X}_{m_1m_2n_1n_2l_1l_2}(3)=-\tfrac{1}{5}[(\delta^{m_1m_2}_{n_1n_2}\Sigma_{l_1l_2}+\delta^{n_1n_2}_{l_1l_2}\Sigma_{m_1m_2}$$

$$+\delta^{l_1l_2}_{m_1m_2}\Sigma_{n_1n_2})+(\delta^{m_1m_2}_{[l_1n_1}\Sigma_{n_2l_2]}$$

$$+\delta^{n_1n_2}_{[m_1l_1}\Sigma_{l_2m_2]}+\delta^{l_1l_2}_{[m_1n_1}\Sigma_{n_2m_2]})], \tag{3.10b}$$

and with increasing r, the $\mathbf{X}(r)$ become increasingly cumbersome in this form.

It is easy to verify that for the BPST field configurations (1.2), for which the curvature $F_{\mu\nu}$ in $4r$ dimensions is proportional to the $2^{2r-1}\times 2^{2r-1}$ SO$(4r)$ spinor matrix $\Sigma_{\mu\nu}$, (3.9) and (3.10) imply that $F(2r)=\hat{F}(2r)$. Thus as far as the BPST solutions are concerned, our modification of the GYM system is of no consequence. Otherwise stated, the GYM fields on $4p$-dimensional spheres are insensitive to the above modification of the dynamics.

Similarly, for the axially symmetric GYM field configurations[5] in eight dimensions, where $x_8=\tau$ has been singled out and the corresponding "radial" variable is then

$r=[x_1^2+\cdots+x_7^2]^{1/2}$, it again turns out that $F(4)=\widehat{F}(4)$.

This situation does not persist, however, and already for the axial symmetry characterized by $\tau=\sqrt{x_6^2+x_7^2+x_8^2}$ and $r=\sqrt{x_1^2+\cdots+x_5^2}$ in eight dimensions, $F(4)\neq\widehat{F}(4)$. Such field configurations have not been considered in Ref. 5, and we intend to examine them elsewhere.

Concerning the properties of the modGYM systems under compactification,[7,8] it turns out that if the compactification is made over a sphere S^n, then the residual system of the modGYM reduces to the usual YMH system in the limit where the radius of compactification is very small, namely, at low energy. Thus even though it has different dynamics, it shares this physically important property with the GYM.

It is easy to demonstrate this property for compactification with spheres, but since we are not concerned in this paper with any solutions—compactifying or otherwise—this qualitative remark shall suffice. We expect that these qualitative properties will persist when spheres are replaced by other coset spaces.

Whether or not the dynamics of modGYM (3.5), as opposed to that of the GYM system (3.2), can lead to any physical consequences is not known. Here we have introduced the modified GYM for our calculational convenience in Sec. IV, and since their relation to GYM is given so simply by (3.9), one can systematically translate our results of Sec. IV to the GYM case.

B. The stress tensor

The stress tensor Θ_a^μ appearing in the Einstein equation is the tensor-valued field of the interacting GEC–GYM system, in addition to the tensor field $E_a^\mu=e^{-1}t_a^\mu$ [cf. Eq. (2.2a)].

Here we are interested in finding that Θ_a^μ vanishes whenever the duality equation (3.1) or (3.6) is satisfied, according to whether gravity is interacting with a GYM or a modified GYM system, respectively.

Since all subsequent manipulations are entirely independent of whether we use $F(2r)$ or $\widehat{F}(2r)$, we shall restrict ourselves to $F(2r)$ only, namely, the GYM system. Our conclusions remain true for the modGYM provided the duality relation (3.1) is replaced by (3.6). Using the metric instead of the *Vielbein*, the stress tensor for the system (3.2) is

$$\Theta_{\mu\nu}=4p\left[\operatorname{tr}F_{\mu\mu_2\cdots\mu_{2p}}F_\nu^{\ \mu_2\cdots\mu_{2p}}-(1/4p)g_{\mu\nu}\operatorname{tr}F^2_{\mu_1\cdots\mu_{2p}}\right]+(i\kappa^2)^{2(q-p)}(2p!/2q!)4q\left[\operatorname{tr}F_{\mu\mu_2\cdots\mu_{2q}}F_\nu^{\ \mu_2\cdots\mu_{2q}}-(1/2q)g_{\mu\nu}\operatorname{tr}F^2_{\mu_1\cdots\mu_{2q}}\right]. \qquad (3.11)$$

We note that $\Theta_\mu^\mu=0$ only when $q=p$, and then (3.2) is a conformally invariant system.

By repeated use of the definition of the dual ${}^*F(2p)$ of $F(2q)$ [given by (1.1b)] and its inverse, the expression (3.11) for $\Theta_{\mu\nu}$ can be recast into the form

$$\Theta_{\mu\nu}=2p\operatorname{tr}(F_{\mu\mu_2\cdots\mu_{2p}}F_{\nu\mu_2\cdots\mu_{2p}}-{}^*F_{\mu\mu_2\cdots\mu_{2p}}{}^*F_{\nu\mu_2\cdots\mu_{2p}})+(i\kappa^2)^{2(q-p)}(2p!/2q!)2p\operatorname{tr}(F_{\mu\mu_2\cdots\mu_{2q}}F_{\nu\mu_2\cdots\mu_{2q}}-{}^*F_{\mu\mu_2\cdots\mu_{2q}}{}^*F_{\nu\mu_2\cdots\mu_{2q}}). \qquad (3.12)$$

This expression for $\Theta_{\mu\nu}$ vanishes when the duality relations (3.1) are satisfied. As stated above, (3.12) holds also with $F(2p)$ and $F(2q)$ replaced by $\widehat{F}(2p)$ and $\widehat{F}(2q)$, namely, for the modGYM system.

IV. SPIN-CONNECTION FIELDS

Here we will use the results of the previous two sections to formulate the main result of this paper, namely, GYM "spin-connection" fields on GEC backgrounds. After that we summarize our results and give a brief discussion.

We consider, in every dimension $d=2(p+q)$, the systems

$$L=L_{\rm GEC}(p,q)+\widehat{L}_{\rm GYM}(p,q), \qquad (4.1)$$

where $L_{\rm GEC}(p,q)$ is given by (2.24), and $\widehat{L}_{\rm GYM}(p,q)$, given by (3.5), is the modGYM system. Subsequently, we shall discuss the replacement of $\widehat{L}_{\rm GYM}(p,q)$ by $L_{\rm GYM}(p,q)$ of (3.2).

In the absence of torsion there will be only one gravitational field equation, the Einstein equation,

$$E_a^\mu=\Theta_a^\mu. \qquad (4.2)$$

The modGYM field equation, on the other hand, is satisfied if the duality relations (3.6) hold. For this (self-dual) modGYM field configuration, the stress tensor Θ_a^μ vanishes identically, as seen from (3.12). Therefore (4.2) reduces to the free Einstein equation of the GEC system (2.24), which we know is satisfied by a (Riemann) double-self-dual field configuration obeying the duality relation (2.16).

There remains then to show that the double-duality relation (2.16) implies the single-duality relation (3.6) when the following identification of the modGYM SO(2$(p+q)$) connection A_μ with the GEC spin connection ω_μ^{ab} is made:

$$A_\mu=-\tfrac{1}{2}\omega_\mu^{mn}\Sigma_{mn}, \qquad (4.3a)$$

$$F_{\mu\nu}=-\tfrac{1}{2}R_{\ \mu\nu}^{mn}\Sigma_{mn}. \qquad (4.3b)$$

In this case, the self-dual modGYM system with gauge connection given by (4.3), interacting with the double-self-dual GEC system with spin connection ω_μ^{mn}, satisfies the Euler–Lagrange equations of the system (4.1). For $p=q=1$, this is precisely the construction that Charap and Duff[27] exploited for the EH–YM system in four dimensions, which we generalize to all even dimensions $d=2(p+q)$. (It should be emphasized that this assertion, that Riemann double-self-duality implies guage single-self-duality, is *not* invertible.)

The proof of this assertion is immediate if the system interacting with gravity is the modGYM, as is the case here in (4.1). In fact this was the reason for introducing the modGYM system in Sec. III A above.

In this case, direct substitution of the double-duality relation (2.16) into the defining expression for $\widehat{F}(2p)$ given by (3.3) and with $F(2)$ given by (4.3b), and noticing that $[(2r-1)(2n-3)\cdots 3]r!=2r!/2^r$, simply yields the duality relation (3.6).

When $\widehat{L}_{\rm GYM}(p,q)$ in (4.1) is replaced by $L_{\rm GYM}(p,q)$ of (3.2), the situation is more complicated. The above assertion, that (Riemann) double-self-duality (2.16) implies GYM single-self-duality (3.1), has to be altered in this case.

For our purposes of constructing spin-connection gauge fields, it is necessary that the GYM duality condition (3.1) hold. This can be achieved by invoking a *supplementary* condition on the Riemann $2p$-form $R(2p)$, *in addition* to the double-duality condition (2.16).

Repeating the above argument we gave in the modGYM case, we see from (3.9) that this supplementary condition is

$$\langle \mathrm{X}(p), R(2p)\rangle = \langle \mathrm{X}(q), {}^*R(2p)\rangle, \tag{4.4a}$$

$$\langle \mathrm{X}(p), R(2p)\rangle \stackrel{\mathrm{def}}{=} \mathrm{X}_{m_1\cdots m_{2p}} R^{\,m_1\cdots m_{2p}}_{\,\mu_1\cdots\mu_{2p}}, \tag{4.4b}$$

$$ {}^*R(2p)^{\,n_1\cdots n_{2q}}_{\,\mu_1\cdots\mu_{2p}} = (e/2q!)(\kappa^2)^{p-q}\varepsilon_{\mu_1\cdots\mu_{2p}\nu_1\cdots\nu_{2q}} R^{\,\nu_1\cdots\nu_{2q}}_{\,n_1\cdots n_{2q}}, \tag{4.4c}$$

or, equivalently, its inverse

$$\langle \mathrm{X}(q), R(2q)\rangle = \langle \mathrm{X}(p), {}^*R(2q)\rangle. \tag{4.5}$$

The expressions $\mathrm{X}(p)$ and $\mathrm{X}(q)$ in (4.4) and (4.5), which are defined by (3.8b), are $2^{p+q-1}\times 2^{p+q-1}$ matrices constructed from the elements of $SO(2(q+p))$ in the spinor representation.

The supplementary duality condition (4.4) or (4.5) is satisfied for the usual metrics on the spheres, as we remarked in Sec. III when we discussed the structure of $\mathrm{X}(r)$ in (3.10). We have checked that (4.4) is satisfied when $p=q=2$, for the HP^2 solution of Ref. 22, and presumably this situation persists when $p=q$, for the CP^{2p} and HP^p metrics.[30] In general, however, the conditions (4.4) become increasingly severe restrictions (with increasing $p+q$) on the otherwise double-self-dual GEC fields, and this may present some obstacles in constructing spin-connection GYM fields, as opposed to the modGYM case.

The main result of this paper is that $SO(2(p+q))$ spin-connection gauge fields (4.3) in $2(p+q)$ dimensions are solutions to the modGYM system (3.5) by virtue of satisfying the modGYM self-duality constraint (3.6), provided that the GEC system (2.24) interacting with it satisfies the double-duality constraint (2.16). An example of such a field configuration is the HP^2 solution[22] for $p=q=2$, which is a member of the hierarchy HP^p for all $p=q$.

The main reason for having introduced the modGYM system in Sec. III was that the above construction can be carried out in exactly the same way for all (p,q), generalizing the $p=q=1$ Einstein–Yang–Mills case considered by Charap and Duff.[27] With gauge group $SO(2(p+q))$, for which they are defined, these modGYM systems are closely related to the GYM systems (3.2), as seen from (3.9). Nevertheless, the dynamics of these two systems given by (3.5) and (3.2), respectively, are quite different, and it is conceivable that for physical reasons one or the other of them may be preferred.

The same construction can still be carried out for spin-connection GYM fields on the GEC background (2.24), provided, however, that the background GEC field configuration satisfies an additional duality condition (4.4). These conditions become increasingly strong constraints in higher dimensions, when the gauge group is also larger. Nevertheless, in the example[22] for the HP^2 ($p=q=2$) spin connection, this additional constraint is satisfied. We do not know whether (4.4) would present an insurmountable obstacle in any given case of interest, but this seems unlikely.

Concerning the *inhomogeneous* double-duality conditions (2.16) when $q>p$, the only solutions we know are the connections on the spheres $S^{2(p+q)}$. On the other hand, several interesting solutions to some of the associated *inhomogeneous* GYM duality equations, especially those with $p=1$, $q>1$, were found by Bais and Batenburg.[14] On the whole, therefore, we can expect that the constructions we propose are not without some interesting examples.

How useful these would prove to be for compactification solutions is an interesting question that we intend to pursue. Such solutions of (2.16) may be of particular interest if we remember that the cosmological constant η, cf. Eq. (2.23), does not necessarily imply a constant curvature space in this case. This last statement is true even for vanishing torsion, while in the presence of torsion this problem was studied in some detail elsewhere.[21]

Our motivation for seeking these spin-connection fields was that of establishing a relationship between the relevant gravitational and gauge field systems in higher dimensions. In the usual KK context, the higher-dimensional system consists of pure gravity, and the YM field appears in the residual system after dimensional reduction. Here, having found (see Appendix) that the GYM systems cannot appear in the dimensional reduction of gravity with a KK *Ansatz*, we have explored the present relationship between the GEC and GYM systems. As opposed to when a KK *Ansatz* is used, here the GEC and GYM systems are treated on the same footing, except that the gravitational double-duality constraint still plays a more fundamental role than the GYM single duality.

In the process, we have put the GYM systems into context, in the background of other work[15,17,18] done in the extension of the YM model to higher dimensions.

ACKNOWLEDGMENTS

We are deeply indepted to A. Chakrabarti for his involvement in this work, at all its stages. One of us (D. H. Tch.) thanks R. Coquereaux and B. P. Dolan for interesting discussions.

APPENDIX: KALUZA–KLEIN REDUCTION OF A GEC SYSTEM

Here we compute the residual GEC system of $L_{\mathrm{GEC}}(p,q)$ with $q=2$, using a KK *Ansatz*. Since we will need all the components of the Riemann tensor, we find it useful to employ the KK reduction formulas of Jensen.[31] We denote the independent components of the higher-dimensional Riemann tensor $\mathbb{R}_{MNRS}=(\mathbb{R}_{\mu\nu\rho\sigma}, \mathbb{R}_{\mu\nu\rho a}, \mathbb{R}_{\mu\nu ab}, \mathbb{R}_{\mu a\nu b}, \mathbb{R}_{ab\mu c}, \mathbb{R}_{abcd})$, with $\mu,\nu,\ldots$ labeling the coordinates of the residual dimensions, and $a,b,\ldots$ labeling the compactified dimensions. These are given by

$$\mathbb{R}_{\mu\nu\rho\sigma} = R_{\mu\nu\rho\sigma} - 2(F^a_{\mu\nu}F^a_{\rho\sigma} + F^a_{\mu\rho}F^a_{\nu\sigma} - F^a_{\nu\rho}F^a_{\mu\sigma}), \tag{A1a}$$

$$\mathbb{R}_{\mu\nu ab} = F^a_{\rho[\mu}F^b_{\nu]\rho} - C^c_{ab}F^c_{\mu\nu}, \tag{A1b}$$

$$\mathbb{R}_{\mu a\nu b} = -F^a_{\mu\rho}F^b_{\rho\nu} + \tfrac{1}{2}F^c_{\mu\nu}C^c_{ab}, \tag{A1c}$$

$$\mathbb{R}_{\mu\nu\rho a} = -D_\rho F^a_{\mu\nu}, \tag{A1d}$$

$$\mathbb{R}_{ab\mu c} = 0, \tag{A1e}$$

$$\mathbb{R}_{abcd} = \tfrac{1}{4} C^{a}_{\ be} C^{e}_{\ cd}. \qquad \text{(A1f)}$$

In (A1), the $F^{a}_{\mu\nu}$ are the YM curvatures in the algebra whose elements are labeled by the index a, D_μ is the covariant derivative with respect to the connection A^{a}_{μ} of this curvature, and the $C^{c}_{\ ab}$ are the structure constants of the Lie group with respect to which the isometries are imposed on the higher-dimensional metric g_{MN}.

Substituting (A1a)–(A1f) into the GEC system (2.1a) with $q = 2$, we find the following residual system:

$$\begin{aligned}
L = {} & (R^2 - 4R_{\mu\nu}R^{\mu\nu} + R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}) \\
& + \tfrac{1}{16}(C^{a}_{\ bc}C^{c}_{\ ab} + C^{d}_{\ ef}C^{f}_{\ de} - 4C^{a}_{\ cd}C^{d}_{\ bc}C^{a}_{\ ef}C^{f}_{\ be} \\
& + C^{a}_{\ be}C^{e}_{\ cd}C^{a}_{\ bf}C^{f}_{\ cd}) + 2R(-\tfrac{1}{4}C^{a}_{\ bc}C^{c}_{\ ab} - F^{a}_{\mu\nu}F^{a}_{\mu\nu}) \\
& + 16R^{\mu\nu}F^{a}_{\mu\rho}F^{a}_{\nu\rho} - \tfrac{1}{2}C^{b}_{\ cd}C^{d}_{\ bc}F^{a}_{\mu\nu}F^{a}_{\mu\nu} \\
& - (2C^{a}_{\ cd}C^{d}_{\ bc} - 3C^{a}_{\ cd}C^{b}_{\ cd})F^{a}_{\mu\nu}F^{b}_{\mu\nu} \\
& + 4(D_\rho F^{a}_{\mu\nu}D_\rho F^{a}_{\mu\nu} - 2D_\nu F^{a}_{\mu\nu}D_\rho F^{a}_{\mu\rho}) \\
& + 12F^{a}_{\mu\rho}F^{b}_{\rho\nu}F^{c}_{\nu\mu}C^{c}_{\ ab} \\
& + [(F^{a}_{\mu\nu}F^{a}_{\mu\nu})^2 - 8(F^{a}_{\mu\rho}F^{a}_{\nu\rho})^2 + 2(F^{a}_{\mu\nu}F^{a}_{\rho\sigma})^2 \\
& + 2F^{a}_{\mu\nu}F^{a}_{\rho\sigma}F^{b}_{\mu\rho}F^{b}_{\nu\sigma}]. \qquad \text{(A2)}
\end{aligned}$$

The first term in (A2) is recognized as $L_{\rm GEC}(p,2)$, and if the dimensionality of the residual system is four, i.e., if $p = 0$, it becomes a pure divergence. The second term is a (cosmological) constant. The fifth and sixth terms give rise to the YM system, while the seventh term would lead to a conformally noninvariant YM field system.

The most interesting for us are the eighth and ninth terms. The former is precisely the first member of the hierarchy of systems proposed by Saçlioğlu,[17] namely, the one that has a BPST instanton in six dimensions. The latter is that one which has the same dimensions (of length) as the $F(4)^2$ GYM system. But even for the simplest gauge group SU(2), this term is *not* equal to the $p = 2$ GYM system.

Our conclusion is that a KK *Ansatz* relates the GEC and GYM systems *only for $p=q=1$, namely, for the Einstein–YM systems only.*

[1] D. H. Tchrakian, J. Math. Phys. **21**, 166 (1980).
[2] D. H. Tchrakian, Phys. Lett. B **150**, 360 (1985).
[3] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin, Phys. Lett. B **59**, 85 (1975).
[4] D. O'Sé and D. H. Tchrakian, Lett. Math. Phys. **13**, 211 (1987).
[5] A. Chakrabarti, T. N. Sherry, and D. H. Tchrakian, Phys. Lett. B **162**, 340 (1985).
[6] E. Witten, Phys. Rev. Lett. **38**, 121 (1977).
[7] P. Forgács and N. S. Manton, Commun. Math. Phys. **72**, 15 (1980).
[8] A. S. Schwarz and Yu. S. Tyupkin, Nucl. Phys. **187**, 321 (1981).
[9] D. O'Sé, T. N. Sherry, and D. H. Tchrakian, J. Math. Phys. **27**, 325 (1986); Zh. Q. Ma, G. M. O'Brien, and D. H. Tchrakian, Phys. Rev. D **33**, 1177 (1986).
[10] D. H. Tchrakian, Phys. Lett. B **155**, 255 (1985).
[11] G. M. O'Brien and D. H. Tchrakian, Nuovo Cimento A **97**, 673 (1987).
[12] B. Grossman, T. W. Kephart, and J. D. Stasheff, Commun. Math. Phys. **96**, 431 (1984); **100**, 311 (1985).
[13] Y. Brihaye, C. Devchand, and J. Nuyts, Phys. Rev. D **32**, 990 (1985).
[14] F. A. Bais and P. Batenburg, Nucl. Phys. B **269**, 363 (1986).
[15] W. Zakrzewski, Nucl. Phys. B **262**, 120 (1985).
[16] B. P. Dolan and D. H. Tchrakian, Phys. Lett. B **198**, 447 (1987).
[17] C. Saçlioğlu, Nucl. Phys. B **277**, 487 (1986).
[18] K. Fujii, Lett. Math. Phys. **12**, 363 (1986).
[19] A. Salam and J. Strathdee, Ann. Phys. (NY) **141**, 316 (1982).
[20] Z. Horvath, L. Palla, E. Cremmer, and J. Scherk, Nucl. Phys. B **127**, 57 (1977); Z. Horvath and L. Palla, *ibid.* **142**, 327 (1978).
[21] D. H. Tchrakian, Gen. Relativ. Gravit. **19**, 135 (1987); Class. Quant. Grav. **4**, L217 (1987).
[22] A. Chakrabarti and D. H. Tchrakian, Phys. Lett. B **168**, 187 (1986).
[23] D. Lovelock, J. Math. Phys. **12**, 498 (1971).
[24] B. Zwiebach, Phys. Lett. B **156**, 315 (1985); E. S. Fradkin and A. A. Tseytlin, Phys. Lett. B **158**, 316 (1985); Nucl. Phys. B **261**, 1 (1985).
[25] A. A. Tseytlin, Phys. Lett. B **176**, 92 (1986); Nucl. Phys. B **276**, 391 (1986).
[26] N. Deruelle and J. Madore, Phys. Lett. A **114**, 185 (1986).
[27] J. M. Charap and M. J. Duff, Phys. Lett. B **69**, 445 (1977).
[28] T. Eguchi and A. J. Hanson, Ann. Phys. (NY) **120**, 82 (1979).
[29] G. M. O'Brien and D.H. Tchrakian, Lett. Math. Phys. **11**, 133 (1986).
[30] N. J. Hitchin (private communication).
[31] G. R. Jensen, J. Differential Geom. **8**, 599 (1973).

Explicit calculation of a model for diffusion in nonconstant temperature

N. G. van Kampen
Institute for Theoretical Physics of the University, Utrecht, The Netherlands

(Received 28 October 1987; accepted for publication 4 November 1987)

As a model for diffusion in a nonhomogeneous medium, Landauer [Phys. Lett. A **68**, 15 (1978)] proposed a pipe filled with a Knudsen gas. The wall temperature varies along the pipe, and the gas molecules, on colliding with the wall, assume its temperature. Here, by explicit calculation, in the limit of small diameter, an equation is derived for the flow of the gas through the pipe. The derivation is possible because the transverse motion of the gas particles rapidly establishes thermal equilibrium. It can therefore be eliminated using the method of eliminating fast variables. The surviving equation for the slow longitudinal flow has the form of a diffusion equation with nonconstant coefficients.

I. INTRODUCTION

The diffusion of a particle in a homogeneous medium at constant temperature, in the presence of an external potential $V(\mathbf{r})$, is described by

$$\frac{\partial P(\mathbf{r},t)}{\partial t} = \nabla\cdot[(\mu\nabla V)P + D\nabla P]. \tag{1}$$

P is the probability density of the particle; the probability flux is minus the quantity in brackets. The mobility μ and diffusion constant D are properties of the medium and its temperature. Throughout we only consider classical particles whose equilibrium distribution is $\exp[-V/kT]$. In order that this be a solution of (1), one must have

$$D = \mu kT. \tag{2}$$

Now suppose that the medium is *not* homogeneous and the temperature is *not* constant in space, so that $\mu(\mathbf{r})$ and $D(\mathbf{r})$ are functions of $\mathbf{r}$. The question of the proper generalization of (1) has been the subject of some debate.[1–6] It is clear that (1) as it stands cannot be valid for nonhomogeneous T; for, if $V=0$, the stationary solution of (1) is $P=$ const, which is at variance with the known phenomenon of thermal diffusion. For this reason some authors proposed

$$\partial_t P = \nabla\cdot[(\mu\nabla V)P + \nabla DP]. \tag{3}$$

Note that this may also be written

$$\partial_t P = \nabla\cdot[(\mu\nabla V + \nabla D)P + D\nabla P], \tag{4}$$

which shows that the difference with (1) consists of an addition to the drift term.

In order to investigate this question, a number of special models have been studied.[1–3,6–9] The purpose of the present article is to work out a model proposed by Landauer.[10] It consists of a particle, or a Knudsen gas, in a thin pipe of constant cross section but whose wall temperature varies along the pipe. On colliding with the wall, the particles assume the local temperature. This is a well-defined model system and the question is how the particle density in the pipe evolves with time. For a constant but infinitely small pipe diameter the model is one dimensional, and only the temperature is not homogeneous. It turns out that the density $P(x,t)$ obeys the diffusion equation

$$\frac{\partial P(x,t)}{\partial t} = \frac{\partial}{\partial x}\left[\mu(x)\frac{dV}{dx}P(x,t) + \frac{\partial}{\partial x}D(x)P(x)\right]. \tag{5}$$

This equation has the form of (3). However, μ and D, apart from being related by (2), have explicit values. In particular, $\mu \propto T^{-1/2}$ and $D \propto T^{1/2}$. The stationary solution of (5) is therefore

$$P^s(x) = \frac{\text{const}}{\sqrt{T(x)}}\exp\left[-\int_0^x \frac{V'(x')}{kT(x')}dx'\right]. \tag{6}$$

The exponential has some interesting consequences.[6,7,11,12] The prefactor $T^{-1/2}$ can be explained by the fact that in the high temperature regions the particle moves faster, the mean velocity being proportional to $T^{1/2}$.[6] A further discussion of this and other results will appear elsewhere.[13] Here the emphasis is on the mathematical formulation of the problem and its solution by means of singular perturbation theory.

II. SPECIFICATION OF THE MODEL

Our system consists of a straight pipe of circular cross section with inner diameter 2ϵ. The axis serves as x axis and the wall is maintained at a temperature $T(x)$. The pipe is filled with a Knudsen gas of particles with coordinates (x,y,z), velocity components (u,v,w), and unit mass. They are subject to an external force with potential $V(x)$, so that their distribution in phase space obeys

$$\frac{\partial f(x,y,z;u,v,w;t)}{\partial t} = -u\frac{\partial f}{\partial x} - v\frac{\partial f}{\partial y} - w\frac{\partial f}{\partial z} + V'(x)\frac{\partial f}{\partial u}. \tag{7}$$

This equation holds for $y^2+z^2<\epsilon^2$. As a boundary condition we assume that a particle, on hitting the inner wall of the pipe, is instantaneously reemitted with a random velocity, whose probability distribution is (the outgoing half of) the Maxwellian corresponding to the local $T(x)$. Clearly, this has the effect that, in the limit $\epsilon\to 0$, the gas itself has the temperature $T(x)$. In the next order of ϵ there will be a deviation from this thermal equilibrium, which makes it possible for particles to drift along the pipe. Our aim is to derive an equation for this drift.

It is convenient to start by formulating the above boundary condition for the case of a half-space $z>0$, with a flat wall at $z=0$. At the wall, the velocity distribution of the emerging particles, $w>0$, is

$$f(x,y,0;u,v,w>0) = A\exp[-\tfrac{1}{2}\beta(u^2+v^2+w^2)],$$

 0022-2488/88/051220-05$02.50

where $\beta(x) = 1/kT(x)$. The factor A is determined by the requirement that the emerging flux must be equal to the incident flux:

$$\int_0^\infty w'\,dw' \int_{-\infty}^\infty du'\,dv'\,f(x,y,0;u',v',w'>0)$$
$$= \int_{-\infty}^0 |w'|dw' \int_{-\infty}^\infty du'\,dv'\,f(x,y,0;u',v',w'<0).$$

Hence the boundary condition reads explicitly

$$\begin{aligned} &f(x,y,0;u,v,w>0)\\ &\quad= (\beta^2/2\pi)\exp[-\tfrac{1}{2}\beta(u^2+v^2+w^2)]\\ &\qquad\times\int_{-\infty}^0 |w'|dw' \int_{-\infty}^\infty du'\,dv'\\ &\qquad\times f(x,y,0;u',v',w'<0). \end{aligned} \tag{8}$$

It expresses the f at positive values of the normal velocity in terms of f at negative values of the normal velocity. Of course, it conserves the total probability: using (7) and partial integration, one has (still for the flat geometry)

$$\frac{d}{dt}\int_0^\infty dz \int_{-\infty}^\infty dx\,dy \int_{-\infty}^\infty du\,dv\,dw\,f(x,y,z;u,v,w;t)$$
$$= \int_{-\infty}^\infty dx\,dy \int_{-\infty}^\infty du\,dv \left[\int_0^\infty w\,dw\,f(x,y,0;u,v,w>0;t) - \int_{-\infty}^0 |w|dw\,f(x,y,0;u,v,w<0;t)\right],$$

which vanishes on account of (8).

Next we adapt our equations to the geometry of the pipe by introducing the cylindrical coordinates

$$y = r\cos\vartheta,\quad z = r\sin\vartheta,$$

and the velocity components in the directions r and ϑ,

$$p = v\cos\vartheta + w\sin\vartheta,\quad rq = -v\sin\vartheta + w\cos\vartheta.$$

The new probability density is

$$F(x,r,\vartheta;u,p,q;t) = r^2 f(x,y,z;u,v,w;t). \tag{9}$$

It obeys the transformed equation

$$\frac{\partial F}{\partial t} = -u\frac{\partial F}{\partial x} - p\frac{\partial F}{\partial r} - q\frac{\partial F}{\partial\vartheta} - rq^2\frac{\partial F}{\partial p} + 2\frac{p}{r}\frac{\partial}{\partial q}qF + V'(x)\frac{\partial F}{\partial u}. \tag{10}$$

This holds for $0 \leqslant r < \epsilon$; the boundary condition is

$$\begin{aligned} &F(x,\epsilon,\vartheta;u,p<0,q)\\ &\quad= (\epsilon\beta^2/2\pi)\exp[-\tfrac{1}{2}\beta(u^2+p^2+\epsilon^2q^2)]\\ &\qquad\times\int_{-\infty}^\infty du'dq' \int_0^\infty p'\,dp'\,F(x,\epsilon,\vartheta;u',p'>0,q'). \end{aligned} \tag{11}$$

Moreover, at $r=0$, one has $F=0$ according to (9), or more precisely,

$$\int_0 F\frac{dr}{r} < \infty. \tag{12}$$

The small parameter ϵ does not appear in the equation but in the geometry of the boundary. We therefore rescale:

$$r = \epsilon\rho,\quad q = \epsilon^{-1}\kappa.$$

As our problem is obviously symmetrical about the axis of the pipe, we save writing by omitting the variable ϑ. Then (10) takes the form

$$\frac{\partial F(x,\rho;u,p,\kappa;t)}{\partial t} = \left[\frac{1}{\epsilon}\mathscr{L}_0 + \mathscr{L}_1\right]F, \tag{13}$$

with two linear operators

$$\mathscr{L}_0 = -p\frac{\partial}{\partial\rho} - \rho\kappa^2\frac{\partial}{\partial p} + 2\frac{p}{\rho}\frac{\partial}{\partial\kappa}\kappa,$$
$$\mathscr{L}_1 = -u\frac{\partial}{\partial\kappa} + V'(x)\frac{\partial}{\partial u}.$$

This holds for $0 \leqslant \rho < 1$; the boundary condition (11) becomes

$$\begin{aligned} &F(x,1;u,p<0,\kappa)\\ &\quad= (\beta^2/2\pi)\exp[-\tfrac{1}{2}\beta(u^2+p^2+\kappa^2)]\\ &\qquad\times\int_{-\infty}^\infty du'\,d\kappa' \int_0^\infty p'\,dp'\,F(x,1;u',p'>0,\kappa'), \end{aligned} \tag{14}$$

and does not involve ϵ. The linear space of functions F obeying (14) and (12) will be denoted by $\mathbb{F}$.

Equation (13) expresses the rate of change of F as the sum of a large term and a term of order unity. The fast change is generated by the operator $\epsilon^{-1}\mathscr{L}_0$, which acts on the transverse coordinate ρ and the transverse velocities p, κ. As a result, the transverse motion of the particles rapidly reaches thermal equilibrium with the wall temperature, namely, in a time of order ϵ. The operator $\mathscr{L}_1$ constitutes a correction, which causes a slow longitudinal drift on the time scale of order ϵ^0. In order to obtain an equation for this drift, we have to eliminate the rapid motion according to the method of eliminating fast variables.[14,15]

III. CONSTRUCTION OF A PROJECTION OPERATOR

For the purpose of eliminating the fast motion, it is necessary to construct a projection operator $\mathscr{P}$ in the space $\mathbb{F}$ with the property

$$\mathscr{P}\mathscr{L}_0 = 0, \tag{15}$$

and preferably also

$$\mathscr{L}_0\mathscr{P} = 0. \tag{16}$$

We start with the latter condition (although it is not the essential one, but it is easier to handle). It states that $\mathscr{P}$ must project $\mathbb{F}$ onto the right null space of $\mathscr{L}_0$. The null space consists of the functions $\Phi\in\mathbb{F}$ obeying

$$\mathscr{L}_0\Phi \equiv -p\frac{\partial\Phi}{\partial\rho} - \rho\kappa^2\frac{\partial\Phi}{\partial p} + 2\frac{p}{\rho}\frac{\partial}{\partial\kappa}\kappa\Phi = 0. \tag{17}$$

The general solution of this equation is

$$\Phi(x,\rho;u,p,\kappa) = \rho^2\Omega(x,u,\rho^2\kappa,(p^2+\rho^2\kappa^2)/2), \tag{18}$$

where Ω is an arbitrary function of its four arguments. Substitution in (14) yields, on setting $\frac{1}{2}(p^2+\kappa^2) = \xi$,

$$\Omega(x,u,\kappa,\xi) = (\beta^2/2\pi)\exp(-\tfrac{1}{2}\beta u^2 - \beta\xi)\times\int_{-\infty}^\infty du'\,d\kappa' \int_{\kappa'^2/2}^\infty d\xi'\,\Omega(x;u',\kappa',\xi').$$

It follows that Ω must have the form

$$\Omega(x,u,\kappa,\xi) = \omega(x)\exp(-\tfrac{1}{2}\beta u^2 - \beta\xi),$$

with arbitrary function ω of x alone. Hence the right null space of $\mathscr{L}_0$ consists of all functions of the form

$$\Phi(x,\rho;u,p,\kappa) = \rho^2\omega(x)\exp\left[-\tfrac{1}{2}\beta(u^2 + p^2 + \rho^2\kappa^2) \right].$$

If one selects successively for $\omega(x)$ the members of a complete orthonormal set $\omega_k(x)$ [for instance, $\omega_k(x) = e^{ikx}/\sqrt{2\pi}$], one obtains an orthogonal frame that spans the right null space.

In order to obey (15) we determine the left null space of $\mathscr{L}_0$, that is, all functions $\Psi(x,\rho;u,p,\kappa)$ that have the property that

$$\int_{-\infty}^{\infty} dx \int_0^1 d\rho\, du\, dp\, d\kappa\, \Psi \mathscr{L}_0 F$$

vanishes for any $F \in \mathbf{F}$. By partial integration one obtains

$$p\frac{\partial\Psi}{\partial\rho} + \rho\kappa^2 \frac{\partial\Psi}{\partial p} - 2\frac{p\kappa}{\rho}\frac{\partial\Psi}{\partial\kappa} = 0, \tag{19}$$

together with the requirement that, for all $F \in \mathbf{F}$,

$$\int_{-\infty}^{\infty} dx\, du\, dp\, d\kappa\, p[\Psi(x,1;u,p,\kappa)F(x,1;u,p,\kappa) - \Psi(x,0;u,p,\kappa)F(x,0;u,p,\kappa)] = 0.$$

The second line of this last equation vanishes if we impose on Ψ the condition that it is finite at $r=0$ [see (12)].

The remaining first line may be written, using (14),

$$\int_{-\infty}^{\infty} dx\, du\, d\kappa \left[\int_0^{\infty} p\, dp\, \Psi(x,1;u,p>0,\kappa)F(x,1;u,p>0,\kappa) - \frac{\beta^2}{2\pi}\int_{-\infty}^{0} |p|dp\, \Psi(x,1;u,p<0,\kappa) \times\exp\left[-\tfrac{1}{2}\beta(u^2 + p^2 + \kappa^2)\right] \times\int_{-\infty}^{\infty} du'\, d\kappa' \int_0^{\infty} p'\, dp'\, F(x,1;u',p'>0,\kappa') \right].$$

This vanishes for any choice of $F(x,1;u',p'>0,\kappa')$ if we impose on Ψ the boundary condition

$$\Psi(x,1;u,p>0,\kappa) = \frac{\beta^2}{2\pi}\int_{-\infty}^{\infty} du'\, d\kappa' \int_{-\infty}^{0} |p'|dp' \times\exp\left[-\tfrac{1}{2}\beta(u'^2 + p'^2 + \kappa'^2)\right]\Psi(x,1;u',p'<0,\kappa'). \tag{20}$$

The functions obeying this condition constitute a dual function space $\mathbf{F}^+$.

The general solution of (19) is

$$\Psi(x,\rho;u,p,\kappa) = \Omega(x,u,\rho^2\kappa,\tfrac{1}{2}(p^2 + \rho^2\kappa^2)),$$

where Ω is again an arbitrary function of four arguments. Substitution in (20) shows that Ω cannot depend on anything except x, so that the left null space is made up of all functions $\omega(x)$. It is again convenient to select a complete orthonormal set $\omega_k^*(x)$, namely, the conjugates of the $\omega_k(x)$ that spanned the right null space.

The projection operator $\mathscr{P}$ can now be constructed[15] simply by setting

$$\mathscr{P}(x,\rho;u,p,\kappa|x',\rho;u',p',\kappa') = 2\left(\frac{\beta}{2\pi}\right)^{3/2}\rho^2 \exp\left[-\tfrac{1}{2}\beta(u^2 + p^2 + \rho^2\kappa^2)\right] \times\sum_k \omega_k(x)\omega_k^*(x') = 2\left(\frac{\beta}{2\pi}\right)^{3/2}\rho^2 \times\exp\left[-\tfrac{1}{2}\beta(u^2 + p^2 + \rho^2\kappa^2)\right]\delta(x-x').$$

It manifestly obeys (15) and (16). Its action on any $F \in \mathbf{F}$ is to integrate out the variables ρ,u,p,κ:

$$\mathscr{P}F(x,\rho;u,p,\kappa) = 2\left(\frac{\beta}{2\pi}\right)^{3/2}\rho^2 \exp\left[-\tfrac{1}{2}\beta(u^2 + p^2 + \rho^2\kappa^2)\right]P(x),$$

where

$$P(x) = \int_0^1 d\rho' \int_{-\infty}^{\infty} du'\, dp'\, d\kappa'\, F(x',\rho';u',p',\kappa') \tag{21}$$

is the density distribution along the pipe, for which we seek an equation.

IV. ELIMINATION OF THE FAST MOTION

On applying $\mathscr{P}$ to (13) one obtains

$$\partial_t \mathscr{P}F = \mathscr{P}\mathscr{L}^{(1)}\mathscr{P}F + \mathscr{P}L^{(1)}\mathscr{Q}F,$$

where $\mathscr{Q} = 1 - \mathscr{P}$. This shows that $\mathscr{P}F$ varies on the slow time scale, but it is not a closed equation. The general theory[15] does lead to a closed equation for $\mathscr{P}F$ in the form of an (asymptotic) expansion in ϵ:

$$\partial_t \mathscr{P}F = \mathscr{P}\mathscr{L}_1\mathscr{P}F - \epsilon\mathscr{P}\mathscr{L}_1\mathscr{Q}\mathscr{L}_0^{-1}\mathscr{Q}\mathscr{L}_1\mathscr{P}F + \mathscr{O}(\epsilon^2). \tag{22}$$

All we have to do is to work out the operators.

It is readily seen that $\mathscr{P}\mathscr{L}_1\mathscr{P} = 0$. The next term in (22), after canceling out a common factor, gives

$$\frac{\partial P(x,t)}{\partial t} = -\epsilon\int_0^1 d\rho \int_{-\infty}^{\infty} du\, dp\, d\kappa \times\mathscr{L}_1\mathscr{Q}\mathscr{L}_0^{-1}\mathscr{Q}\mathscr{L}_1\mathscr{P}F.$$

The two factors $\mathscr{Q}$ take away the null vectors of $\mathscr{L}_0$ and thereby guarantee that $\mathscr{L}_0^{-1}$ exists and is unique. In the present context they may be omitted because we know that $\mathscr{Q}\mathscr{L}^{(1)}\mathscr{P} = \mathscr{L}^{(1)}\mathscr{P}$ and $\mathscr{P}\mathscr{L}^{(1)}\mathscr{Q} = \mathscr{P}\mathscr{L}^{(1)}$. On substituting $\mathscr{L}_1$ one obtains

$$\frac{\partial P(x,t)}{\partial t} = \epsilon\frac{\partial}{\partial x}\int_{-\infty}^{\infty} u\, du\, dp\, d\kappa \int_0^1 d\rho\mathscr{L}_0^{-1} \times\left(-u\frac{\partial}{\partial x} + V'(x)\frac{\partial}{\partial u}\right) 2\left(\frac{\beta}{2\pi}\right)^{3/2}\rho^2 \times\exp\left[-\tfrac{1}{2}\beta(u^2 + p^2 + \rho^2\kappa^2)\right]P(x,t). \tag{23}$$

Evidently the right-hand side contains a first and a second derivative of $P(x,t)$ with respect to x, and has therefore the form of a diffusion equation. First consider the term involving $V'(x)$:

$$I = \frac{\partial}{\partial x}\left[2\epsilon\left(\frac{\beta}{2\pi}\right)^{3/2}(-\beta)\int_{-\infty}^{\infty} u\,du\,dp\,d\kappa\int_0^1 d\rho \times \mathscr{L}_0^{-1}\rho^2 u\exp\left(-\frac{1}{2}\beta(u^2+p^2+\rho^2\kappa^2)\right)\right] \times V'(x)P(x,t). \tag{24}$$

Clearly, the expression in brackets is the mobility μ, which depends on x through the temperature $\beta(x)$.

In order to evaluate the integral we have to find

$$G(x,\rho;u,p,\kappa) = \mathscr{L}_0^{-1}\rho^2 u\exp[-\tfrac{1}{2}\beta(u^2+p^2+\rho^2\kappa^2)]. \tag{25}$$

That is, we have to solve G from

$$\mathscr{L}_0 G \equiv -p\frac{\partial G}{\partial \rho} - \rho\kappa^2\frac{\partial G}{\partial p} + 2\frac{p}{\rho}\frac{\partial}{\partial\kappa}\kappa G = \rho^2 u\exp[-\tfrac{1}{2}\beta(u^2+p^2+\rho^2\kappa^2)]. \tag{26}$$

The general solution is found to be

$$G = \rho^2 u\exp[-\tfrac{1}{2}\beta(u^2+p^2+\rho^2\kappa^2)] \times\left[\frac{-\rho p}{p^2+\rho^2\kappa^2} + \Omega\left(x,u,\rho^2\kappa,\frac{p^2+\rho^2\kappa^2}{2}\right)\right],$$

where Ω is again an arbitrary function. In order that G lie in F, it must obey (14), which tells us that

$$\frac{-p}{p^2+\kappa^2} + \Omega\left(x,u,\kappa,\frac{p^2+\kappa^2}{2}\right) = \frac{\omega(x)}{u} \quad (p<0), \tag{27}$$

where $\omega(x)$ stands for

$$\omega(x) = \frac{\beta^2}{2\pi}\int_{-\infty}^{\infty} du'\,d\kappa'\int_0^{\infty} p'\,dp'\,G(x,1;u',p',\kappa'),$$

and may be any function of x.

The function Ω of its arguments x, u, η, ξ is determined by (27):

$$\Omega(x,u,\eta,\xi) = \frac{\omega(x)}{u} + \frac{p}{p^2+\kappa^2} = \frac{\omega(x)}{u} - \frac{\sqrt{2\xi-\eta^2}}{2\xi}.$$

Hence (25) becomes

$$G(x,\rho;u,p,\kappa) = \rho^2\exp[-\tfrac{1}{2}\beta(u^2+p^2+\rho^2\kappa^2)] \times\left[\frac{-\rho u p}{p^2+\rho^2\kappa^2} + \omega(x) - u\frac{\sqrt{p^2+\rho^2\kappa^2-\rho^4\kappa^2}}{p^2+\rho^2\kappa^2}\right]. \tag{28}$$

From this we need only the part that survives after applying $\mathscr{Q} = 1 - \mathscr{P}$, which amounts to omitting $\omega(x)$.

On substituting this result in (24) one obtains

$$I = \frac{\partial}{\partial x}\mu(x)V'(x)P(x,t),$$

$$\mu(x) = -2\epsilon\frac{\beta}{2\pi}\int_{-\infty}^{\infty} dp\,d\kappa\int_0^1 \rho^2\,d\rho \times\exp[-\tfrac{1}{2}\beta(p^2+\rho^2\kappa^2)] \times\left[-\frac{\rho p}{p^2+\rho^2\kappa^2} - \frac{\sqrt{p^2+\rho^2\kappa^2-\rho^4\kappa^2}}{p^2+\rho^2\kappa^2}\right].$$

The first term in the brackets gives zero contribution, and the second one can be evaluated by using the polar coordinates $p = s\cos\varphi$, $\rho\kappa = s\sin\varphi$:

$$\mu = \frac{\epsilon\beta}{\pi}\int_0^{\infty}\rho\,d\rho\int_0^{\infty}\exp\left(-\frac{1}{2}\beta s^2\right)s\,ds\int_0^{2\pi}d\varphi \times\frac{\sqrt{1-\rho^2\sin^2\varphi}}{s} = \frac{8}{3}\epsilon\sqrt{\frac{\beta}{2\pi}}. \tag{29}$$

Thus we have found

$$I = \frac{8\epsilon}{3\sqrt{2\pi}}\frac{\partial}{\partial x}\beta^{1/2}V'(x)P(x,t). \tag{30}$$

V. THE DIFFUSION EQUATION

Having computed the term of (23) that involves $V'(x)$, we now turn to the remaining one. It may be written as the sum of two contributions, viz.,

$$II = -2\epsilon\frac{\partial}{\partial x}\int_{-\infty}^{\infty} u\,du\,dp\,d\kappa\int_0^1 d\rho\,\mathscr{L}_0^{-1} \times\rho^2 u\exp[-\tfrac{1}{2}\beta(u^2+p^2+\rho^2\kappa^2)] \times\frac{\partial}{\partial x}\left(\frac{\beta}{2\pi}\right)^{3/2}P,$$

$$III = -2\epsilon\frac{\partial}{\partial x}\int_{-\infty}^{\infty} u\,du\,dp\,d\kappa\int_0^1 d\rho\,\mathscr{L}_0^{-1} \times\left[\frac{\partial}{\partial x}\rho^2 u\exp\left(-\frac{1}{2}\beta(u^2+p^2+\rho^2\kappa^2)\right)\right] \times\left(\frac{\beta}{2\pi}\right)^{1/2}P.$$

The integral in II is the same as in (24), so that without further calculation we find

$$II = \frac{\partial}{\partial x}\mu\beta^{-5/2}\frac{\partial}{\partial x}\beta^{3/2}P = \frac{8\epsilon}{3\sqrt{2\pi}}\frac{\partial}{\partial x}\beta^{-2}\frac{\partial}{\partial x}\beta^{3/2}P. \tag{31}$$

To evaluate III we need to solve H from

$$\mathscr{L}_0 H = \frac{\partial}{\partial x}\rho^2 u\exp[-\tfrac{1}{2}\beta(u^2+p^2+\rho^2\kappa^2)]. \tag{32}$$

A special solution can be obtained directly by applying $\partial/\partial x$ to the special solution (28) of (26):

$$H = \frac{1}{2}\rho^2\frac{\rho p+\sqrt{p^2+\rho^2\kappa^2-\rho^4\kappa^2}}{p^2+\rho^2\kappa^2} \times\frac{d\beta}{dx}u(u^2+p^2+\rho^2\kappa^2) \times\exp[-\tfrac{1}{2}\beta(u^2+p^2+\rho^2\kappa^2)].$$

It obeys the boundary condition (14) and also $\mathscr{P}H = 0$. Hence this is the solution of (32) we need, so that

$$III = -2\epsilon\frac{\partial}{\partial x}\int_{-\infty}^{\infty} u\,du\,dp\,d\kappa\int_0^1 d\rho\,H\left[\frac{\beta}{2\pi}\right]^{3/2}P = -\epsilon\frac{\partial}{\partial x}\left[\frac{\beta}{2\pi}\right]^{3/2}P\frac{d\beta}{dx}\int_0^1\rho^2\,d\rho\int_{-\infty}^{\infty}dp\,d\kappa$$

$$\times\frac{\sqrt{p^2+\rho^2\kappa^2-\rho^4\kappa^2}}{p^2+\rho^2\kappa^2}\sqrt{\frac{2\pi}{\beta}}\left[\frac{3}{\beta^2}+\frac{p^2+\rho^2\kappa^2}{\beta}\right]$$

$$\times\exp[-\tfrac{1}{2}\beta(p^2+\rho^2\kappa^2)]P$$

$$=-\frac{8\epsilon}{3}\frac{\partial}{\partial x}\sqrt{\frac{2\beta}{\pi}}\beta^{-2}\frac{d\beta}{dx}P$$

$$=-\frac{16\epsilon}{3\sqrt{2\pi}}\frac{\partial}{\partial x}\beta^{-3/2}\frac{d\beta}{dx}P. \tag{33}$$

Collecting the results (30), (31), and (33), one has finally

$$\frac{\partial P}{\partial t}=\frac{8\epsilon}{3\sqrt{2\pi}}\frac{\partial}{\partial x}$$

$$\times\left[\beta^{1/2}V'P-\frac{1}{2}\beta^{-3/2}\frac{d\beta}{dx}P+\beta^{-1/2}\frac{\partial P}{\partial x}\right]$$

$$=\frac{8\epsilon}{3\sqrt{2\pi}}\frac{\partial}{\partial x}\left[T^{-1/2}V'P+\frac{\partial}{\partial x}kT^{1/2}P\right]. \tag{34}$$

This result may also be written in the form

$$\frac{\partial P}{\partial t}=\frac{\partial}{\partial x}\left[\mu V'P+\frac{\partial}{\partial x}\mu kTP\right]. \tag{35}$$

Thus we have derived (5) together with (2).

The fact that P does indeed obey a diffusion equation is by no means self-evident. Not only is it confined to the first order of ϵ, but also it ceases to be true if one takes, instead of the three-dimensional pipe, a two-dimensional strip. The physical reason is that in two dimensions there are too many particles moving almost parallel with the walls; they transport mass over a large distance between two randomizing collisions with the wall. Mathematically this shows up as a divergence of the integrals at small transverse velocities. Specifically, in (29) the factor s in the numerator is absent.

[1]R. Stratton, Phys. Rev. **126**, 2002 (1962); J. Appl. Phys. **40**, 4582 (1969); IEEE Trans. Electron Devices **ED-19**, 1288 (1972).

[2]J. D. Gassaway, IEEE Trans. Electron Devices **ED-18**, 175 (1971).

[3]A. H. Marshak and D. Assaf III, Solid-State Electron. **16**, 675 (1973); C. Baccarani and A. M. Mazzone, Solid-State Electron. **18**, 469 (1975).

[4]P. T. Landsberg and S. A. Hope, Solid-State Electron. **19**, 173 (1976); **20**, 421 (1977).

[5]M. Büttiker and R. Landauer, in *Nonlinear Phenomena at Phase Transitions and Instabilities*, edited by T. Riste (Plenum, New York 1982), p. 111.

[6]R. Landauer, Helv. Phys. Acta **56**, 847 (1983).

[7]N. G. van Kampen, IBM J. Res. Dev. (to be published).

[8]N. G. van Kampen, Z. Phys. B **68**, 135 (1987).

[9]N. G. van Kampen, *Proceedings of the International School of Atomic Molecular Spectroscopy VIII*, Erice, Italy, June 1987 (to be published).

[10]R. Landauer, Phys. Lett. A **68**, 15 (1978).

[11]R. Landauer, Phys. Rev. A **12**, 636 (1975).

[12]M. Büttiker, Z. Phys. B **68**, 161 (1987).

[13]N. G. van Kampen, J. Phys. Chem. Solids (to be published).

[14]U. Geigenmüller, U. M. Titulaer, and B. U. Felderhof, Physica A **119**, 41, 53 (1983); **120**, 635 (1983); C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 1983).

[15]N. G. van Kampen, Phys. Rep. **124**, 69 (1985).

On canonical irreducible quantum field theories describing bosons and fermions

Klaus Baumann
Institut für Theoretische Physik, Bunsenstrasse 9, D-3400 Göttingen, West Germany

(Received 16 June 1987; accepted for publication 13 January 1988)

For a relativistic irreducible quantum field theory (QFT) in which the bosons and fermions fulfill the canonical commutation relation (CCR) and the canonical anticommutation relation (CAR), respectively, it is shown that in $n>3$ space dimensions Bose and Fermi fields fulfill free field equations. Furthermore, interactions that might be compatible with CCR and CAR in lower space dimensions are characterized. The possible candidates are interactions of type $Q_2(\psi,\psi^\dagger)\phi + P_4(\phi)$ in $n=3$, $Q_2(\psi,\psi^\dagger)P_2(\phi) + P_6(\phi)$ in $n=2$, and $Q_4(\psi,\psi^\dagger)F(\phi)$ in $n=1$ space dimensions, where Q_m and P_m are polynomials of degree up to m.

I. INTRODUCTION

Powers[1] has shown that a relativistic irreducible Fermi field fulfilling the canonical anticommutation relation (CAR) is a free field in $n>1$ space dimensions. We have supplemented his result by showing that for $n=1$ the interaction can be at most of quartic type.[2] Based on estimates given by Herbst[3] we have demonstrated in Ref. 4 that a relativistic irreducible scalar Bose field fulfilling the canonical commutation relation (CCR) is necessarily a free field in $n>3$ space dimensions. In lower space dimensions we got restrictions on the type of interaction that looks formally like a $P_4(\phi)$ interaction for $n=3$ [resp. a $P_6(\phi)$ interaction for $n=2$ space dimensions].

In this paper we consider a canonical Wightman field theory describing a scalar Bose field and Fermi fields. We show that in $n>3$ space dimensions bosons and fermions fulfill free field equations. In lower space dimensions we get restrictions that look formally like an interaction of type $Q_2(\psi,\psi^\dagger)\phi + P_4(\phi)$ for $n=3$, $Q_2(\psi,\psi^\dagger)P_2(\phi) + P_6(\phi)$ for $n=2$, and $Q_4(\psi,\psi^\dagger)F(\phi)$ for $n=1$ space dimensions. Here Q_m and P_m denote polynomials of degree up to m. In the language of standard perturbation theory this result can be stated as follows: All nonrenormalizable interactions can never fulfill CCR [resp. CAR].

On the other hand constructive quantum field theory has dealt successfully with all super-renormalizable models, e.g., $:P(\phi):_{1+1}$, $:\bar\psi\psi\phi:_{1+1}$, $:\phi^4:_{2+1}$, and $:\bar\psi\psi\phi:_{2+1}$.

For the remaining class of renormalizable interactions like, e.g., $:(\bar\psi\gamma_\mu\psi)(\bar\psi\gamma^\mu\psi):_{1+1}$, $:\phi^6:_{2+1}$, $:\phi^4:_{3+1}$, or $:\bar\psi\psi\phi:_{3+1}$, we cannot exclude the possibility of canonical commutation relations but at least we get bounds for the coupling constants if we keep track of all the constants appearing in our estimates.

The main trick for proving the above result is to decouple the estimates for mixed commutators in such a way that we can use the estimates for the pure CAR and the pure CCR cases. This will be done in Sec. III of this paper. Based on these estimates we shall state and prove our main result in Sec. IV. In Sec. II we formulate our assumptions.

II. ASSUMPTIONS

A. Relativistic quantum field theory

Throughout this paper we make the usual assumptions for a Wightman field theory in $n+1$ space-time dimensions supplemented by the requirement that sharp time fields exist.

The $\psi_k(t,x)$, $k=1,\ldots,m$, together with their adjoints $\psi_k^\dagger(t,x)$ describe an m-component Fermi field and $\phi(t,x)$ describes a neutral, scalar Bose field. Here H denotes the Hamiltonian (generator of time translations). Because the estimates for the CCR case are based on Araki's formula (see Ref. 3)

$$(e^{i\phi(t,f)}\Omega,He^{i\phi(t,g)}\Omega) = \tfrac{1}{2}(f,g)(e^{i\phi(t,f)}\Omega,e^{i\phi(t,g)}\Omega)$$

we make the following assumption.

Consider the subspace $\mathscr{H}_B \subseteq \mathscr{H}$ generated by the linear span of the vectors

$$\{\Omega,\phi(F_1)\Omega,\ldots,\phi(F_k)\cdots\phi(F_1)\Omega,\ldots|F_k\in\mathscr{S}(\mathbb{R}^{n+1}),\ k\in\mathbb{N}\}.$$

We assume that the linear span of the vectors

$$\{\Omega,\phi(0,f_1)\Omega,\ldots,\phi(0,f_k)\times\cdots\phi(0,f_1)\Omega,\ldots|f_k\in\mathscr{S}(\mathbb{R}^n),\ k\in\mathbb{N}\}$$

is a dense set in $\mathscr{H}_B$ and determines the Hamiltonian H restricted to $\mathscr{H}_B$ uniquely.

With $\pi(t,x)$ we denote the canonical momentum associated with $\phi(t,x)$. For convenience let us agree to the following convention: If we write down a field operator without an explicit time argument, we mean always the field operator at time zero, i.e., $\phi(f)\equiv\phi(0,f)$, $f\in\mathscr{S}(\mathbb{R}^n)$, $\pi(f)\equiv\pi(0,f)$ and so on.

B. Commutation relations

The Fermi fields $\psi_k(t,f)$ and $\psi_k(t,f)^\dagger$, $f\in\mathscr{S}(\mathbb{R}^n)$, are bounded operators and fulfill the canonical anticommutation relations (CAR):

$$\{\psi_k(t,f),\psi_l(t,g)\} = 0 = \{\psi_k(t,f)^\dagger,\psi_l(t,g)^\dagger\}, \tag{2.1}$$

$$\{\psi_k(t,f)^\dagger,\psi_l(t,g)\} = \delta_{kl}\int_{\mathbb{R}^n}(\bar f g)(x)\,d^n x. \tag{2.2}$$

To control the unboundedness of the Bose fields $\phi(t,f)$ and $\pi(t,f)$, $f\in\mathscr{S}(\mathbb{R}^n)$, we adopt Fröhlich's formulation of CCR (see Ref. 5).

Definition 2.1: We say an operator C fulfills a form bound relative to $H+1$ if its real part $\operatorname{Re} C = \tfrac{1}{2}(C+C^\dagger)$

 0022-2488/88/051225-06$02.50

and its imaginary part $\operatorname{Im} C = (1/2i)(C - C^\dagger)$ fulfill on the domain $\hat{Q}(H)$ the form bounds

$$\pm \operatorname{Re} C \leqslant \gamma_+ (H+1) \quad \text{and} \quad \pm \operatorname{Im} C \leqslant \gamma_- (H+1), \tag{2.3}$$

where γ_+ and γ_- are positive real numbers depending on C.

Now we can formulate canonical commutation relations (CCR) as follows.

For all f and $g \in \mathscr{S}(\mathbb{R}^n)$ the operators $\phi(t,f)$ and $\pi(t,g)$ fulfill form bounds relative to $H+1$ and have furthermore

$$[\phi(t,f),\phi(t,g)] = 0 = [\pi(t,f),\pi(t,g)], \tag{2.4}$$

$$[\phi(t,f),\pi(t,g)] = i \int_{\mathbb{R}^n} (fg)(x)\, d^n x, \tag{2.5}$$

weakly on $D(H) \times D(H)$.

In addition to CAR and CCR we assume for the mixed commutators

$$[\phi(t,f),\psi_k(t,h)] = 0 = [\phi(t,f),\psi_k(t,h)^\dagger] \tag{2.6}$$

and

$$[\pi(t,g),\psi_k(t,h)] = 0 = [\pi(t,g),\psi_k(t,h)^\dagger] \tag{2.7}$$

for all $f,g,h \in \mathscr{S}(\mathbb{R}^n)$ and $k = 1,\dots,m$.

C. Irreducibility

Normally irreducibility is formulated as follows: A bounded operator B that commutes with $e^{i\phi(0,f)}$ and $e^{i\pi(0,g)}$ for all $f,g \in \mathscr{S}_{\text{real}}(\mathbb{R}^n)$ and also with $\psi(0,h)$ and $\psi(0,h)^\dagger$ for all $h \in \mathscr{S}(\mathbb{R}^n)$ is a C number, i.e., $B = (\Omega, B\Omega)$.

This formulation is not very practical for our purposes and therefore we shall use Fröhlich's commutator theorem[5] to cast irreducibility in the following sufficient form.

Proposition 2.2: Assume the operator C fulfills (a) a form bound relative to $H+1$; (b) C and $C^\dagger$ commute with the time zero fields, i.e.,

$$\begin{aligned}
&[C,\phi(f)] = 0 = [C^\dagger,\phi(f)],\\
&[C,\pi(g)] = 0 = [C^\dagger,\pi(g)],\\
&[C,\psi(h)] = 0 = [C^\dagger,\psi(h)],\\
&[C,\psi(h)^\dagger] = 0 = [C^\dagger,\psi(h)^\dagger];
\end{aligned} \tag{2.8}$$

then C is a C number, i.e., $C = (\Omega, C\Omega)$.

Proof: This is just an application of Fröhlich's commutator theorem (see Ref. 5) combined with "usual" irreducibility as formulated above.

Remark 2.3: If C contains an odd number of Fermi fields and therefore anticommutes with ψ then $C = 0$ (see Ref. 1).

D. Existence of $\dot{\pi}$ and $\dot{\psi}$

We assume that for $g,h \in \mathscr{S}(\mathbb{R}^n)$ the operators $\dot{\pi}(t,g)$, $\dot{\psi}(t,h)$, and $\dot{\psi}(t,h)^\dagger$ exist and fulfill form bounds relative to $H+1$.

Remark 2.3: We think that with a little bit more effort one could weaken this assumption considerably. Powers,[1] for example, assumed only that $\dot{\psi}(t,h)^\# \Omega$ and $\dot{\psi}(t,h)^\# \psi(t,h')^\# \Omega$ exist for all $h,h' \in \mathscr{S}(\mathbb{R}^n)$ [$\psi(t,h)^\#$ denotes either $\psi(t,h)$ or $\psi(t,h)^\dagger$!]. We think that similar assumptions are sufficient in the case of Bose fields, too. But such weaker assumptions would complicate many proofs considerably.

E. Technical assumptions

Because our proof relies heavily on irreducibility we need form bounds relative to $H+1$ for certain multiple commutators like, e.g.,

$$[\pi(g_1),\dot{\pi}(g_0)],[\pi(g_2)[\pi(g_1),\dot{\pi}(g_0)]],\dots .$$

Therefore whenever we use irreducibility we assume implicitly the corresponding form bound relative to $H+1$.

III. ESTIMATES FOR MIXED COMMUTATORS

One of the central points in Powers' work[1] on CAR was to show that in $n > 1$ space dimensions

$$[\psi_{k_2}(h_2)^\# \{\psi_{k_1}(h_1)^\#,\dot{\psi}_{k_0}(h_0)^\#\}] \equiv 0, \tag{3.1}$$

for all $h_i \in \mathscr{S}(\mathbb{R}^n)$ and $k_i \in \{1,\dots,m\}$. For $n=1$ we have shown in Ref. 2 that

$$\Big[\psi_{k_4}(h_4)^\# \{\psi_{k_3}(h_3)^\# [\psi_{k_2}(h_2)^\# \{\psi_{k_1}(h_1)^\#,\dot{\psi}_{k_0}(h_0)^\#\}]\}\Big] \equiv 0, \tag{3.2}$$

for all $h_i \in \mathscr{S}(\mathbb{R})$ and $k_i \in \{1,\dots,m\}$.

For the case of CCR (see Ref. 4) we got the result that

$$[\pi(g_N)[\pi(g_{N-1})[\cdots[\pi(g_1),\dot{\pi}(g_0)]\cdots] \equiv 0, \tag{3.3}$$

for all $g_k \in \mathscr{S}(\mathbb{R}^n)$ if the number n of space dimensions is larger than $(N+3)/(N-1)$, where N is the number of commutators involved in (3.3). From the commutation relation (2.5) we derive immediately

$$[\phi(g_1),\dot{\pi}(g_0)] \equiv 0, \quad \text{for all } g_0,g_1 \in \mathscr{S}(\mathbb{R}^n), \tag{3.4}$$

and by iterative use of the Jacobi identity we get

$$[\phi(g_N)[\pi(g_{N-1})[\cdots[\pi(g_1),\dot{\pi}(g_0)]\cdots] \equiv 0 \tag{3.5}$$

for all $g_i \in \mathscr{S}(\mathbb{R}^n)$ as has been shown in Ref. 4.

In the case of a canonical field theory where we have simultaneously fermions and bosons all these estimates remain true. Therefore the only new estimates we need are those for mixed commutators involving Fermi and Bose fields.

If we differentiate the mixed commutators (2.6) and (2.7) with respect to the time t we get the algebraic relations

$$[\phi(f),\dot{\psi}_k(h)^\#] = 0 \tag{3.6}$$

and

$$[\pi(g),\dot{\psi}_k(h)^\#] = [\psi_k(h)^\#,\dot{\pi}(g)], \tag{3.7}$$

for all $f,g,h \in \mathscr{S}(\mathbb{R}^n)$ on suitable dense domains.

A first step towards a standard form for mixed commutators is given in the following lemma.

Lemma 3.1: Let Φ represent $\phi(f)$ or $\pi(g)$ and Ψ represent $\psi_k(h)$ or $\psi_k(h)^\dagger$; then

$$[\Psi[\Phi,B]] = [\Phi[\Psi,B]] \tag{3.8}$$

and

$$\{\Psi[\Phi,B]\} = [\Phi\{\Psi,B\}]. \tag{3.9}$$

Proof: From the commutators (2.6) and (2.7) it follows

$$\begin{aligned}
[\Psi[\Phi,B]] &= \Psi\Phi B - \Psi B\Phi - \Phi B\Psi + B\Phi\Psi\\
&= \Phi\Psi B - \Psi B\Phi - \Phi B\Psi + B\Psi\Phi\\
&= [\Phi[\Psi,B]],
\end{aligned}$$

and similarly

$$\{\Psi[\Phi,B]\} = \Psi\Phi B - \Psi B\Phi + \Phi B\Psi - B\Phi\Psi$$
$$= \Phi\Psi B - \Psi B\Phi + \Phi B\Psi - B\Psi\Phi$$
$$= [\Phi\{\Psi,B\}].$$

Therefore any multiple commutator under consideration containing M Bose operators Φ and L Fermi operators Ψ can always be written in one of the two equivalent *standard forms*:

$$\underbrace{\cdots\{\Psi[\Psi[}_{L\Psi\text{'s}}\underbrace{\Phi[\Phi\cdots[\Phi,\dot{\Phi}]}_{M\Phi\text{'s}}\cdots\}\cdots \tag{3.10}$$

or equivalently

$$\underbrace{[\Phi[\Phi\cdots[\Phi,}_{M\Phi\text{'s}}\underbrace{\cdots[\Psi\{\Psi,\dot{\Psi}\}]}_{L\Psi\text{'s}}\cdots] . \tag{3.11}$$

[This follows from Lemma 3.1 and relation (3.7).]

As an immediate consequence from the above standard form we get the following.

(i) Any commutator containing a time zero field $\phi(f)$ vanishes. [This follows from (3.4)–(3.6).]

(ii) Any commutator containing more than two Fermi fields vanishes in $n>1$ space dimensions. In one space dimension any commutator containing more than four Fermi fields vanishes. [This follows from (3.1) and (3.2).]

Therefore it is sufficient to consider the following two types of mixed commutators:

$$\text{(a)}\quad [\psi_k(g_N)^{\#}[\pi(g_{N-1})[\cdots[\pi(g_1),\dot{\pi}(g_0)]\cdots] \tag{3.12}$$

and

$$\text{(b)}\quad [\pi(g_N)[\cdots[\pi(g_2)\{\psi_k(g_1)^{\#},\psi_l(g_0)\}]\cdots]. \tag{3.13}$$

In the following we shall show that the estimates already obtained in the pure CAR [resp. CCR] cases are sufficient to estimate the above mixed commutators.

Let us briefly outline on what ideas the following proofs are based.

We shall show that the mixed commutators (3.12) and (3.13) applied on the vacuum state Ω vanish for sufficiently large N. It turns out that N will never exceed 6!

In a first step we approximate the time zero operators $\pi(g)$, $\dot{\pi}(g)$, and $\dot{\psi}(g)^{\#}$ by $(-1)\phi(\dot{f}^\epsilon,g)$, $\phi(\ddot{f}^\epsilon,g)$, and $(-1)\psi(\dot{f}^\epsilon,g)^{\#}$, respectively, where $f^\epsilon\in\mathscr{D}(\mathbf{R})$ is a δ sequence as ϵ goes to zero. The corresponding vectors converge strongly in $\mathscr{H}$.

In a second step we use a smooth partition E_k^ϵ of the unity to chop the test functions g_1 into small pieces $E_k^\epsilon g_1$, each of which occupies a volume of about $(\frac{3}{2}\epsilon)^n$. Due to locality a lot of terms do not contribute.

The estimates given by Herbst[3] for the Bose fields are based on $\nabla\phi$ bounds. Therefore we replace $\phi(\dot{f}^\epsilon,E_k^\epsilon g)$ by $\phi(\dot{f}^\epsilon,E_k^\epsilon g-(E_k^\epsilon g)_{\alpha\epsilon e_1})$ where the subscript $\alpha\epsilon e_1$ means translation by $\alpha\epsilon e_1$. With α suitably chosen these additional terms do not affect the commutators because of locality. But now we can write

$$\phi(\dot{f}^\epsilon,E_k^\epsilon g-(E_k^\epsilon g)_{\alpha\epsilon e_1})\quad\text{as}\quad\phi(\dot{f}^\epsilon,\partial_1 H_k^\epsilon)$$

and it is obvious that the L_2 norms of $\partial_1 H_k^\epsilon$ and H_k^ϵ are bounded as follows:

$$\|\partial_1 H_k^\epsilon\|_2 < C\cdot\max|g|\epsilon^{n/2}$$

and

$$\|H_k^\epsilon\|_2 < \hat{C}\cdot\max|g|\epsilon^{n/2+1}.$$

Because the strong limit of

$$\Psi^\epsilon := \sum[\psi(0,E_{k_N}^\epsilon g_N)[\phi(\dot{f}^\epsilon,E_{k_{N-1}}^\epsilon g_{N-1})$$
$$\times[\cdots,\phi(\ddot{f}^\epsilon,E_{k_0}^\epsilon g_0)]\cdots]\Omega$$

exists it is sufficient to estimate the matrix element $(P(\phi,\Psi)\Omega,\Psi^\epsilon)$, where we choose the set of operators $P(\phi,\Psi)$ in such a way that $P(\phi,\Psi)\Omega$ forms a dense set in $\mathscr{H}$ and furthermore $P(\phi,\Psi)$ commutes with all the operators on the rhs of Ψ^ϵ. By this method we are able to decouple the estimates for Bose and Fermi fields!

The Fermi fields are bounded operators, i.e., $\|\psi(t,h)^{\#}\| \leq \|h\|_2$, and because of this fact each $\psi(0,E_k^\epsilon g_l)$ contributes a factor $\sim\epsilon^{n/2}$, and by the same reasoning $\psi(\dot{f}^\epsilon,E_k^\epsilon g_1)$ contributes a factor $\sim\epsilon^{n/2-1}$. The Bose fields are always unbounded operators, but from the estimates given by Herbst[3] we can conclude that within a Wightman function of finite order each $\phi(0,\partial_1 H_{k,l}^\epsilon)$, $\phi(\dot{f}^\epsilon,\partial_1 H_{k,l}^\epsilon)$, and $\phi(\ddot{f}^\epsilon,\partial_1 H_{k,l}^\epsilon)$ contributes a factor $\sim\epsilon^{(n+1)/2}$, $\sim\epsilon^{(n-1)/2}$, and $\sim\epsilon^{(n-3)/2}$, respectively. The above exponents of ϵ are optimal because even the two-point functions $\|\phi(0,E_k^\epsilon g)\Omega\|\sim\epsilon^{(n+1)/2}$, $\|\pi(0,E_k^\epsilon g)\Omega\|\sim\epsilon^{(n-1)/2}$, and $\|\dot{\pi}(0,E_k^\epsilon g)\Omega\|\sim\epsilon^{(n-3)/2}$ do not behave better. This demonstrates how powerful Herbst's estimates are!

Let us start with multiple commutators containing only one Fermi field.

Lemma 3.2: For $g_0,\ldots,g_N\in\mathscr{D}(\mathbf{R}^n)$ we have

$$[\psi_l(g_N)^{\#}[\pi(g_{N-1})[\cdots[\pi(g_1),\dot{\pi}(g_0)]\cdots]\Omega\equiv 0, \tag{3.14}$$

in $n>(N+2)/(N-1)$ space dimensions.

Proof: (a) As in the case of CCR we approximate

$$[\pi(g_{N-1})[\cdots[\pi(g_1),\dot{\pi}(g_0)]\cdots]$$

by

$$(-1)^{N+1}[\phi(\dot{f}_{N-1}^\epsilon,g_{N-1})$$
$$\times[\cdots[\phi(\dot{f}_1^\epsilon,g_1),\phi(\ddot{f}_0^\epsilon,g_0)]\cdots]$$

with (i) $f_k\in\mathscr{D}([-\frac{1}{10},\frac{1}{10}])$, (ii) $\int f_k(t)dt=1$, and for $\epsilon>0$ we define $f_k^\epsilon(t)=(1/\epsilon)f_k(t/\epsilon)$. Because we work within the Wightman framework it is clear from our assumptions that

$$\lim_{\epsilon\to 0}\|(-1)^{N+1}[\psi_l(g_N)^{\#}[\phi(\dot{f}_{N-1}^\epsilon,g_{N-1})$$
$$\times[\cdots,\phi(\ddot{f}_0^\epsilon,g_0)]\cdots]\Omega$$
$$-[\psi_l(g_N)^{\#}[\pi(g_{N-1})[\cdots,\dot{\pi}(g_0)]\cdots]\Omega\|=0. \tag{3.15}$$

Therefore it is sufficient to show that for all $\Psi\in\mathscr{D}$

$$\lim_{\epsilon\to 0}(\Psi,[\psi_l(g_N)^{\#}[\cdots,\phi(\ddot{f}_0^\epsilon,g_0)]\cdots]\Omega)=0, \tag{3.16}$$

where $\mathscr{D} \subseteq \mathscr{H}$ is a dense set of vectors in the Hilbert space $\mathscr{H}$.

(b) Assume that supp $g_0,\ldots,$supp $g_N \subseteq O \subseteq \mathbf{R}^n$, O compact. Take $\hat{O} \subseteq \mathbf{R}^{n+1}$ a compact, nonempty set with $\hat{O}$ spacelike to $[-1,1] \times O \subseteq \mathbf{R}^{n+1}$. Let $P_L(\phi,\psi,\psi^\dagger)$ be a monomial of degree $L \geqslant 0$ in the field operators smeared with test functions supported in $\hat{O}$. By the Reeh–Schlieder theorem the linear span of $\{P_L(\phi,\psi,\psi^\dagger)\Omega | P_L \in \mathscr{P}(\hat{O})\}$ is a dense set in $\mathscr{H}$ and therefore we put

$$\mathscr{D} = \langle P_L(\phi,\psi,\psi^\dagger)\Omega | P_L \in \mathscr{P}(\hat{O}),\ L \geqslant 0\rangle. \tag{3.17}$$

And because Wightman functions containing an odd number of Fermi fields vanish, we can assume for the following that $P_L(\phi,\psi,\psi^\dagger)$ contains an odd number of Fermi fields.

(c) Let E^ϵ_k, $k \in \mathbf{Z}^n$, be a smooth partition of the unity as defined in Appendix A of Ref. 4. By linearity we have

$$(P_L\Omega,[\psi_l(g_N)^\#[\cdots,\phi(\ddot{f}^\epsilon_0,g_0)]\cdots]\Omega)$$
$$= \sum_{k_0,\ldots,k_N \in \mathbf{Z}^n} (P_L\Omega,[\psi_l(E^\epsilon_{k_N} g_N) \times [\cdots,\phi(\ddot{f}^\epsilon_0,E^\epsilon_{k_0} g_0)]\cdots]\Omega) \tag{3.18}$$

and from locality we get the restrictions

$$|(k_j - k_0)_i| \leqslant j, \quad \text{for } i = 1,2,\ldots,n \text{ and } j = 1,2,\ldots,N. \tag{3.19}$$

This sum has at most $[(2N+1)!!(L+2)/\epsilon]^n$ terms if supp g_0 is contained in $[-L/2,L/2] \times \cdots \times [-L/2, L/2]$.

(d) The decoupling of Fermi and Bose fields is based on the following proposition.

Proposition 3.3: Let A denote

$$[\phi(\dot{f}^\epsilon_{N-1},E^\epsilon_{k_{N-1}} g_{N-1})[\cdots,\phi(\ddot{f}^\epsilon_0,E^\epsilon_{k_0} g_0)]\cdots];$$

then

$$|(P_L\Omega,[\psi_l(E^\epsilon_{k_N} g_N),A\,]\Omega)|$$
$$\leqslant \|\psi_l(E^\epsilon_{k_N} g_N)\| \{\|P_L\Omega\| \cdot \|A^\dagger\Omega\| + \|P^\dagger_L\Omega\| \cdot \|A\Omega\|\}. \tag{3.20}$$

Proof:

$$|(P_L\Omega,[\psi_l(E^\epsilon_{k_N} g_N)^\#,A\Omega)|$$
$$= |(P_L\Omega,\psi_l(E^\epsilon_{k_N} g_N)^\# A\Omega) - (P_L\Omega,A\psi_l(E^\epsilon_{k_N} g_N)^\#\Omega)|$$
$$\leqslant |(P_L\Omega,\psi_l(E^\epsilon_{k_N} g_N)^\# A\Omega)|$$
$$+ |(A^\dagger,\psi_l(E^\epsilon_{k_N} g_N)^\# P^\dagger_L\Omega|$$

because by construction $[P^\dagger_L,A\,] = 0 = \{P^\dagger_L, \psi_l(E^\epsilon_{k_N} g_N)^\#\}$. Proposition 3.3 follows from the Cauchy–Schwarz inequality and the fact that $\psi_l(E^\epsilon_{k_N} g_N)^\#$ is a bounded operator.

(e) For the final estimate we use

$$\|\psi_l(E^\epsilon_{k_N} g_N)^\#\| \leqslant \max|g_N| \cdot \epsilon^{n/2} \tag{3.21}$$

and from our previous work on CCR (see Proposition 2.3 of Ref. 4) we know that

$$\|[\phi(\dot{f}^\epsilon_{N-1},E^\epsilon_{k_{N-1}} g_{N-1})[\cdots,\phi(\ddot{f}^\epsilon_0,E^\epsilon_{k_0} g_0)]\cdots]\Omega\|$$
$$\leqslant C(f_{N-1},\ldots,f_0;g_{N-1},\ldots,g_0)\epsilon^{(1/2)(Nn-N-2)}. \tag{3.22}$$

These bounds imply

$$|(P_L\Omega,[\psi_l(g_N)^\#[\phi(\dot{f}^\epsilon_{N-1},g_{N-1})[\cdots,\phi(\ddot{f}^\epsilon_0,g_0)]\cdots]\Omega)|$$
$$\leqslant \{(2N+1)!![(L+2)/\epsilon]\}^n \max|g_N|\epsilon^{n/2}$$
$$\times \{\|P_L\Omega\| + \|P^\dagger_L\Omega\|\}$$
$$\cdot C(f_{N-1},\ldots,f_0;g_{N-1},\ldots,g_0)\epsilon^{(1/2)(Nn-N-2)}. \tag{3.23}$$

With respect to ϵ this behaves like $\epsilon^{[(N-1)/2](n-(N+2)/N-1))}$. As ϵ goes to zero Lemma 3.2 follows.

Next we consider $[\pi(g_2)\{\psi_k(g_1)^\#,\dot{\psi}_l(g_0)^\#\}]$ which is the simplest case of the multiple commutator containing two Fermi fields. One motivation for dealing first with this special case is that the estimates for the general case containing several are more involved and this might confuse the reader. Also it is sufficient to estimate the above double commutator for showing that in more than three space dimensions $\phi(t,x)$ and $\psi(t,x)$ obey free field equations!

Lemma 3.4: For $n > 3$ and $g_0,g_1,g_2 \in \mathscr{D}(\mathbf{R}^n)$ we have

$$[\pi(g_2)\{\psi_k(g_1)^\#,\dot{\psi}_l(g_0)^\#\}]\Omega \equiv 0. \tag{3.24}$$

Proof: (a) The first steps are the same as in the proof of Lemma 3.2, except that we can now assume that $P_L(\phi,\psi,\psi^\dagger) \in \mathscr{P}(\hat{O})$ contains an even number of Fermi fields. Therefore we have

$$(P_L\Omega,[\pi(g_2)\{\psi_k(g_1)^\#,\dot{\psi}_l(g_0)^\#\}]\Omega)$$
$$= \sum_{k_2,k_1,k_0 \in \mathbf{Z}^n} (P_L\Omega,[\pi(E^\epsilon_{k_2} g_2) \times \{\psi_k(E^\epsilon_{k_1} g_1)^\#,\psi_l(\dot{f}^\epsilon_0,E^\epsilon_{k_0} g_0)^\#\}]\Omega), \tag{3.25}$$

with $|(k_1 - k_0)_i| \leqslant 1$ and $|(k_2 - k_0)_i| \leqslant 2$ for $i = 1,\ldots,n$.

(b) Each term on the rhs of (3.25) can be estimated as follows:

$$|(P_L\Omega,[\pi(E^\epsilon_{k_2} g_2)\{\psi_k(E^\epsilon_{k_1} g_1)^\#,\psi_l(\dot{f}^\epsilon_0,E^\epsilon_{k_0} g_0)^\#\}]\Omega)|$$
$$\leqslant |(\pi(E^\epsilon_{k_2} g_2)^\dagger\Omega,\{\psi_k(E^\epsilon_{k_1} g_1)^\#,\psi_l(\dot{f}^\epsilon_0,E^\epsilon_{k_0} g_0)^\#\}P^\dagger_L\Omega)| + |(P_L\Omega,\{\psi_k(E^\epsilon_{k_1} g_1)^\#,\psi_l(\dot{f}^\epsilon_0,E^\epsilon_{k_0} g_0)^\#\}\pi(E^\epsilon_{k_2} g_2)\Omega)|$$
$$\leqslant 2\|\psi_k(E^\epsilon_{k_1} g_1)^\#\| \cdot \|\psi_l(E^\epsilon_{k_0} g_0)^\#\| \int_{\mathbf{R}} |\dot{f}^\epsilon_0(t)|dt \cdot \{\|P^\dagger_L\Omega\| \cdot \|\pi(E^\epsilon_{k_2} g_2)^\dagger\Omega\| + \|P_L\Omega\| \cdot \|\pi(E^\epsilon_{k_2} g_2)\Omega\|\}. \tag{3.26}$$

Now

$$\int |\dot{f}^\epsilon_0(t)|dt = \frac{1}{\epsilon^2}\int \left|\dot{f}_0\left(\frac{t}{\epsilon}\right)\right| dt = \frac{1}{\epsilon}\int |\dot{f}_0(t)|dt \tag{3.27}$$

and from the Källen–Lehmann representation for the two-point function we get

$$\|\pi(E^{\epsilon}_{k}g)\Omega\|^2=\int_0^\infty dm^2\rho(m^2)\int_{\mathbb{R}^n}dp\frac{\sqrt{m^2+p^2}}{2}|(E^{\epsilon}_{k}g)(p)|^2$$
$$\leqslant\frac{1}{2}\int_0^\infty dm^2\rho(m^2)\left[\int_{\mathbb{R}^n}dp(m^2+p^2)|(E^{\epsilon}_{k}g)(p)|^2\right]^{1/2}\left[\int_{\mathbb{R}^n}dp|(E^{\epsilon}_{k}g)(p)|^2\right]^{1/2}$$
$$=\frac{1}{2}\int_0^\infty dm^2\rho(m^2)[m^2\|E^{\epsilon}_{k}g\|_2^2+\|\nabla E^{\epsilon}_{k}g\|_2^2]^{1/2}\|E^{\epsilon}_{k}g\|_2$$
$$\leqslant\frac{1}{2}\int_0^\infty dm^2\rho(m^2)\{m\|E^{\epsilon}_{k}g\|_2^2+\|\nabla E^{\epsilon}_{k}g\|_2\|E^{\epsilon}_{k}g\|_2\}$$
$$\leqslant(\max|g|+\max|\nabla g|)(C\cdot\epsilon^{n-1}+D\cdot\epsilon^n). \tag{3.28}$$

Therefore the rhs of (3.26) is proportional to $\epsilon^{(3/2)n-3/2}$ and because there are $(3\cdot5)^n((L+2)/\epsilon)^n$ terms we get finally

$$|(P_L\Omega,[\pi(g_2)\{\psi_k(g_1)^\#,\psi_l(\dot{f}^{\epsilon}_0,g_0)\}]\Omega)|\sim\epsilon^{(1/2)(n-3)}. \tag{3.29}$$

As ϵ goes to zero Lemma 3.4 follows. As a curiosity let us remark that for Lemma 3.4 we did not use the estimates given by Herbst[3] because the bound (3.28) could be deduced from the Källen–Lehmann representation.

Finally let us briefly describe how to handle the general case of an N-fold commutator containing two Fermi fields.

Lemma 3.5: For $g_0,\dots,g_N\in\mathscr{D}(\mathbb{R}^n)$ we have

$$[\pi(g_N)[\cdots[\pi(g_2)\{\psi_k(g_1)^\#,\dot{\psi}_l(g_0)\}]\cdots]\Omega\equiv0, \tag{3.30}$$

for $n>(N+1)/(N-1)$ space dimensions.

Proof: (a) Lemma 3.4 covers the case $N=2$ ($\hat{=}\, n>3$ space dimensions). The other N's of interest are $N=3$ ($\hat{=}\, n>2$ space dimensions) and $N=4$ ($\hat{=}\, n>1$ space dimension). For simplicity let us take $N=3$. The first steps are the same as in the proof of Lemma 3.2. We have to estimate

$$(P_L\Omega,[\phi(\dot{f}^{\epsilon}_3,g_3)[\phi(\dot{f}^{\epsilon}_2,g_2)\{\psi_k(g_1)^\#,\psi_l(\dot{f}^{\epsilon}_0,g_0)\}]]\Omega)$$
$$=\sum_{k_3,\dots,k_0\in\mathbb{Z}^n}(P_L\Omega,[\phi(\dot{f}^{\epsilon}_3,E^{\epsilon}_{k_3}g_3)$$
$$\times[\cdots,\psi_l(\dot{f}^{\epsilon}_0,E^{\epsilon}_{k_0}g_0)\}]]\Omega), \tag{3.31}$$

with the restrictions $|(k_j-k_0)_i|\leqslant j$, $i=1,\dots,n$, $j=1,2,3$, implied by locality.

(b) If we replace $E^{\epsilon}_{k_j}g_j$ by $E^{\epsilon}_{k_j}g_j-(E^{\epsilon}_{k_j}g_j)_{g\epsilon e_1}$, where the subscript $g\epsilon e_1$ means translation by $g\epsilon$ in the x_1 direction for $j=2,3$ (for the Bose fields only!), we do not affect the rhs of (3.31) because of locality! Obviously we can write

$$E^{\epsilon}_{k_j}g_j-(E^{\epsilon}_{k_j}g_j)_{(\pm)g\epsilon e_1}=\partial_1H^{\epsilon}_{j,k_j},\quad\text{for } j=2,3,$$

where H^{ϵ}_{j,k_j} are elements of $\mathscr{D}(\mathbb{R}^n)$. This strange looking procedure is necessary because the estimates for the Bose fields given by Herbst[3] are based on $\nabla\phi$ bounds. Instead of (3.31) we get

$$(P_L\Omega,[\phi(\dot{f}^{\epsilon}_3,g_3)[\phi(\dot{f}^{\epsilon}_2,g_2)\{\psi_k(g_1)^\#,\psi_l(\dot{f}^{\epsilon}_0,g_0)^\#\}]]\Omega)$$
$$=\sum_{k_3,\dots,k_0\in\mathbb{Z}^n}(P_L\Omega,[\phi(\dot{f}^{\epsilon}_3,\partial_1H^{\epsilon}_{3,k_3})[\phi(\dot{f}^{\epsilon}_2,\partial_1H^{\epsilon}_{2,k_2})$$
$$\times\{\psi_k(E^{\epsilon}_{k_1}g_1)^\#,\psi_l(\dot{f}^{\epsilon}_0,E^{\epsilon}_{k_0}g_0)\}]]\Omega), \tag{3.32}$$

with $|(k_j-k_0)_i|\leqslant j$ for $i=1,\dots,n$ and $j=1,2,3$.

(c) Now we decouple Bose and Fermi operators similar to Proposition 3.3 as follows [we use $\phi(3)$, $\phi(2)$, $\psi(1)$, and $\psi(0)$ as a shorthand notation for $\phi(\dot{f}^{\epsilon}_3,\partial_1H^{\epsilon}_{3,k_3}),\dots,\psi_l(\dot{f}^{\epsilon}_0,E^{\epsilon}_{k_0}g_0)^\#!$]:

$$|(P_L\Omega,[\phi(3)[\phi(2)\{\psi(1),\psi(0)\}]]\Omega)|$$
$$\leqslant\|\{\psi(1),\psi(0)\}\|(\|\phi(2)^\dagger\phi(3)^\dagger\Omega\|\cdot\|P^\dagger_L\Omega\|$$
$$+\|P_L\phi(3)^\dagger\Omega\|\,\|\phi(2)\Omega\|+\|P_L\phi(2)^\dagger\Omega\|\,\|\phi(3)\Omega\|$$
$$+\|\phi(2)\phi(3)\Omega\|\cdot\|P_L\Omega\|). \tag{3.33}$$

Furthermore for the second and third term we have

$$\|P_L\phi(3)^\dagger\Omega\|=|(\Omega,\phi(3)P^\dagger_LP_L\phi(3)^\dagger\Omega)|^{1/2}$$
$$=|(P^\dagger_LP_L\Omega,\phi(3)\phi(3)^\dagger\Omega)|^{1/2} \tag{3.34}$$

for P_L and $P^\dagger_L$ commute with $\phi(3)$ by construction, and from the Cauchy–Schwarz inequality get

$$\leqslant\|P^\dagger_LP_L\Omega\|^{1/2}\|\phi(3)\phi(3)^\dagger\Omega\|^{1/2}. \tag{3.35}$$

From (3.33) together with (3.35) we see that for the Bose field operators we have to estimate configurations of the type

$$\|\phi(\dot{f}^{\epsilon}_3,\partial_1H^{\epsilon}_{3,k_3})\phi(\dot{f}^{\epsilon}_2,\partial_1H^{\epsilon}_{2,k_2})\Omega\| \tag{3.36}$$

as ϵ goes to zero. (For $N=4$ we need estimates for the six-point functions too!) Therefore we can use the estimates given by Herbst (see Refs. 3 and 4) for the pure CCR case. From these estimates we know that (3.36) behaves like $\sim\epsilon^{n-1}$. Therefore we have

$$|(P_L\Omega,[\phi(\dot{f}^{\epsilon}_3,g_3)[\cdots,\psi_l(\dot{f}^{\epsilon}_0,g_0)\}]]\Omega)|\sim\epsilon^{n-2}. \tag{3.37}$$

As ϵ goes to zero Lemma 3.5 for the special case $N=3$ follows.

IV. ON THE TRIVIALITY OF ϕ AND ψ

Now we use the results from the previous section to show that in $n>3$ space dimensions ϕ and ψ obey free field equations and how the type of interaction is restricted in lower space dimensions.

Lemma 4.1: For $n>3$ and $f\in\mathscr{D}(\mathbb{R}^n)$, $g\in\mathscr{S}(\mathbb{R}^n)$ we have

$$[\pi(g),\dot{\psi}_k(f)^\#]=0=[\psi_k(g)^\#,\dot{\pi}(f)]. \tag{4.1}$$

Proof: From the Reeh–Schlieder theorem we know that Lemmas 3.2 and 3.5 imply

$$[\psi_k(g_N)^\#[\pi(g_{N-1})[\cdots,\dot{\pi}(g_0)]\cdots]\equiv0,$$
$$\text{for } n>(N+2)/(N-1) \tag{4.2}$$

and

$$[\pi(g_N)[\cdots[\pi(g_2)\{\dot\psi_k(g_1)^{\#},\dot\psi_l(g_0)^{\#}\}\cdots]\equiv 0,\quad \text{for } n>(N+1)/(N-1). \tag{4.3}$$

Using the algebraic identities of Sec. III we have for $n>3$

$$[\phi(h_2)[\pi(h_1)[\pi(g),\dot\psi_k(f)^{\#}]]]=0, \tag{4.4a}$$

$$[\pi(h_2)[\pi(h_1)[\pi(g),\dot\psi_k(f)^{\#}]]]=0, \tag{4.4b}$$

$$\{\psi_l(h_2)^{\#}[\pi(h_1)[\pi(g),\dot\psi_k(f)^{\#}]]\}=0. \tag{4.4c}$$

These relations remain true even if $g,h_1,h_2\in\mathscr{S}(\mathbb{R}^n)$ because as long as f has compact support locality acts as an effective cutoff! In our technical assumptions (5) we have assumed a form bound for $[\pi(h_1)[\pi(g),\dot\psi_k(f)^{\#}]]$ and from irreducibility we get

$$[\pi(h_1)[\pi(g),\dot\psi_k(f)^{\#}]] = (\Omega,[\pi(h_1)[\pi(g),\dot\psi_k(f)^{\#}]]\Omega)=0 \tag{4.5}$$

because the vacuum expectation value contains only one fermion. Therefore we have for $n>3$

$$[\phi(h)[\pi(g),\dot\psi_k(f)^{\#}]]=0, \tag{4.6a}$$

$$[\pi(h)[\pi(g),\dot\psi_k(f)^{\#}]]=0, \tag{4.6b}$$

$$\{\psi_l(h)^{\#}[\pi(g),\dot\psi_k(f)^{\#}]\}=0. \tag{4.6c}$$

Again we conclude from irreducibility that

$$[\pi(g),\dot\psi_k(f)^{\#}] = (\Omega,[\pi(g),\dot\psi_k(f)^{\#}]\Omega)=0. \tag{4.7}$$

The relation $[\psi_k(g)^{\#},\dot\pi(f)]=0$ follows from (3.7).

Now we can state our main result in the following theorem.

Theorem 4.2: In $n>3$ space dimensions $\phi(t,x)$ and $\psi(t,x)$ fulfill free field equations as given in the previous papers.[2,4]

Proof: (a) From the pure Bose case,[4] from (3.5), and from Lemma 3.2 we know that for $n>3$,

$$[\phi(h)[\pi(g),\dot\pi(f)]]=0, \tag{4.8a}$$

$$[\pi(h)[\pi(g),\dot\pi(f)]]=0, \tag{4.8b}$$

$$[\psi_k(h)^{\#}[\pi(g),\dot\pi(f)]]=0. \tag{4.8c}$$

From irreducibility we conclude that

$$[\pi(g),\dot\pi(f)]=(\Omega,[\pi(g),\dot\pi(f)]\Omega). \tag{4.9}$$

But as shown in Ref. 4 from the Källen–Lehmann representation, CCR, and Eq. (4.9) we get

$$[\pi(g),\dot\pi(f)-\phi(\Delta f)+M^2\phi(f)]=0 \tag{4.10}$$

with

$$M^2=\int_0^\infty dm^2\,\rho(m^2)m^2<\infty. \tag{4.11}$$

Furthermore we know that

$$[\phi(g),\dot\pi(f)-\phi(\Delta f)+M^2\phi(f)]=0 \tag{4.10'}$$

and from Lemma 4.1 and (2.6) we get

$$[\psi_k(g)^{\#},\dot\pi(f)-\phi(\Delta f)+M^2\phi(f)]=0. \tag{4.10''}$$

Therefore we can use irreducibility again to conclude for $f\in\mathscr{D}(\mathbb{R}^n)$ that

$$\dot\pi(f)-\phi(\Delta f)+M^2\phi(f)=0 \tag{4.12}$$

because we can assume $(\Omega,\phi(t,x)\Omega)\equiv 0$. By continuity (4.12) remains true for $f\in\mathscr{S}(\mathbb{R}^n)$.

(b) In a similar way we show that

$$\{\psi_k(g)^{\#},\dot\psi_l(f)^{\#}\}=(\Omega,\{\psi_k(g)^{\#},\dot\psi_l(f)^{\#}\}\Omega). \tag{4.13}$$

Using our former results concerning the pure CAR case (see Ref. 2) we get from Lemma 4.1 that $\psi(t,x)$ and $\psi^\dagger(t,x)$ obey a linear first-order partial differential equation. This proves Theorem 4.2.

The following theorem restricts the type of interaction that might be compatible with the assumed commutation relations in less than four space dimensions.

Theorem 4.3: The following commutators are c numbers.

(a) For $n=3$ space dimensions,

$$[\pi(g_2)\{\psi_k(g_1)^{\#},\dot\psi_l(g_0)^{\#}\}] = (\Omega,[\pi(g_2)\{\cdots,\dot\psi_l(g_0)^{\#}\}]\Omega) \tag{4.14}$$

and

$$[\pi(g_3)[\pi(g_2)[\pi(g_1),\dot\pi(g_0)]]] = (\Omega,[\pi(g_3)[\pi(g_2)[\pi(g_1),\dot\pi(g_0)]]]\Omega)$$

"$Q_2(\psi,\psi^\dagger)\phi+P_4(\phi)$ interaction." (4.15)

(b) For $n=2$ space dimensions,

$$[\pi(g_3)[\pi(g_2)\{\psi_k(g_1)^{\#},\dot\psi_l(g_0)^{\#}\}]] = (\Omega,[\pi(g_3)[\cdots,\dot\psi_l(g_0)^{\#}\}]]\Omega) \tag{4.16}$$

and

$$[\pi(g_5)[\cdots[\pi(g_1),\dot\pi(g_0)]\cdots] = (\Omega,[\pi(g_5)[\cdots,\dot\pi(g_0)]\cdots]\Omega)$$

"$Q_2(\psi,\psi^\dagger)P_2(\phi)+P_6(\phi)$ interaction." (4.17)

(c) For $n=1$ space dimension

$$[\psi(g_4)^{\#}\{\psi(g_3)^{\#}[\psi(g_2)^{\#}\{\psi(g_1)^{\#},\psi(g_0)^{\#}\}]\}]\equiv 0$$

"$Q_4(\psi,\psi^\dagger)F(\phi)$ interaction." (4.18)

Proof: (a) and (b) are immediate consequences from Lemma 3.2 and Lemma 3.5 combined with irreducibility. (c) follows already from the pure CAR case (see Ref. 2).

ACKNOWLEDGMENTS

I would like to thank Arthur Wightman for suggesting this problem and for his constant encouragement.

Also many thanks to the Max Kade Foundation, New York, for the financial support of this work in its earlier stages at Princeton University.

[1] R. T. Powers, "Absence of interaction as a consequence of good ultraviolet behavior in the case of a local Fermi field," Commun. Math. Phys. **4**, 145 (1967).

[2] K. Baumann, "Three remarks on Powers' theorem about irreducible fields fulfilling CAR," J. Math. Phys. **27**, 2373 (1986).

[3] I. W. Herbst, "On canonical quantum field theories," J. Math. Phys. **17**, 1210 (1976).

[4] K. Baumann, "On relativistic irreducible quantum fields fulfilling CCR," J. Math. Phys. **28**, 697 (1987).

[5] J. Fröhlich, "Application of commutator theorems to the integration of representations of Lie algebras and commutation relations," Commun. Math. Phys. **54**, 135 (1977).

Krein structures for Wightman and Schwinger functions

L. Jakobczyk[a] and F. Strocchi
International School for Advanced Studies, Trieste, Italy

(Received 25 August 1987; accepted for publication 11 November 1987)

For quantum field theories that do not satisfy the Wightman positivity condition, a Hilbert space structure condition is proposed, which guarantees a Krein structure for the space of states associated to the Wightman functions. The analogous problem for the Schwinger function is also discussed, as well as the conditions that ensure the analytic continuation of a Krein structure in the Euclidean case and vice versa. The Gupta–Bleuler formulation of free quantum electrodynamics is discussed as an example.

I. INTRODUCTION

The relevance of quantum field theories satisfying the Wightman axioms with the exception of positivity[1] is supported by the success of the treatment of gauge field theories in local (*renormalizable*) gauges (as it is done, e.g., in the perturbative approach) and more generally by the interest of theories with infrared singularities preventing a regular behavior of the space-time translations.[2]

A modified version of Wightman axioms with a weak form of the spectral condition and a Hilbert space structure condition replacing the positivity axiom has been discussed in Refs. 2 and 3 and extensively used in the construction of "charged" states in quantum electrodynamics (QED) and in the analysis of their properties,[4,5] as well as in the discussion of two-dimensional quantum field theory models.[6]

The Euclidean formulation of local quantum field theories without positivity has been discussed in Ref. 7; modified Euclidean axioms were presented allowing a recovering of Wightman functions (from the Schwinger functions) satisfying the modified Wightman axioms.

In this paper, we will discuss a stronger form of the Hilbert space structure condition which guarantees the existence of a majorizing Hilbert space structure (associated to the Wightman functions) with the property of being of Krein type.[2] This means that there exists a metric operator η such that $\eta^2 = 1$, and $\eta\Psi_0 = \Psi_0$. Briefly, we will call this stronger form the *Krein structure condition*. A special stronger version of this condition was discussed at the algebraic level in Ref. 8. Furthermore we will discuss the Krein structure condition both at the level of Wightman functions and at the level of Schwinger functions and characterize the condition that allows the analytic continuation of one structure in the other. In this way we will also get a characterization of the Euclidean formulation of quantum field theories with a Krein structure. The above structures are explicitly checked and worked out in the simple case of the Gupta–Bleuler formulation of free QED.

II. KREIN POSITIVITY OF WIGHTMAN AND SCHWINGER FUNCTIONS

A. Relativistic case

We start with the set of Wightman functions $\{W_n\}$ satisfying all Wightman axioms except that of positivity. For simplicity we discuss the case of an Hermitian scalar field (the generalization being straightforward).

W1 *(temperedness):* For any n, W_n is a tempered distribution.

W2 *(Poincaré convariance):* For every $(a,\Lambda)\in P^{\uparrow}_{+}$ and for any n,

$$W_n(x_1,...,x_n) = W_n(\Lambda x_1 - a,...,\Lambda x_n - a)\,.$$

W4 *(weak spectral condition):* For any n,

$$W^d_{n-1}(x_2 - x_1,...,x_n - x_{n-1}) \equiv W_n(x_1,...,x_n)$$

has a Fourier transform $\tilde{W}^d_{n-1}$ with support in $\bar{V}^{n-1}_{+}$.

W5 *(locality):* For any n,

$$W_n(x_1,...,x_i,x_{i+1},...,x_n) = W_n(x_1,...,x_{i+1},x_i,...,x_n)$$

whenever $(x_i - x_{i+1})^2 < 0$.

For the discussion of these axioms, see Refs. 2 and 7. In particular we refer to Ref. 7 for notational details.

We say that the Wightman functions $\{W_n\}$ are Krein positive if[9] they satisfy the following condition.

W3' *(Krein positivity):* There exists a mapping α of some dense subalgebra with identity $\mathscr{B}_0$, of the Borchers algebra[10] $\mathscr{B}$, into itself, such that[11] $\forall F,G\in\mathscr{B}_0$

(1) $W(\{\alpha(\alpha(F))\}^*\times G) = W(F^*\times G)$

$$\equiv \sum_{n,m} W_{n+m}(f^*_n\times g_m);$$

(2) $W(\alpha(F)^*\times F)\geqslant 0$;

(3) $W(\alpha(F)^*\times G) = W(F^*\times\alpha(G))$;

(4) $p_\alpha(F)\equiv W(\alpha(F)^*\times F)^{1/2}$ is continuous in the topology of $\mathscr{B}$ (briefly $\mathscr{B}$-continuous).[7]

More pedantically we should say that in this case the Wightman functions are Krein positive with reference to $\mathscr{B}_0$.

Remark: By $\mathscr{B}$ continuity, p_α can be extended from $\mathscr{B}_0$ to $\mathscr{B}$ (the extension will be denoted by the same symbol).

Proposition 2.1: The seminorm p_α is nondegenerate, i.e.,

$$\ker p_\alpha = \mathscr{N}_W \equiv \{F\in\mathscr{B}:\ W(F^*\times G) = 0,\ \forall G\in\mathscr{B}\}\,.$$

Proof: We first show that p_α defines a majorant topology. By condition (2) of W3' we have

$$|W(\alpha(F)^*\times G)|^2 \leqslant W(\alpha(F)^*\times F)W(\alpha(G)^*\times G),\quad \forall F,G\in\mathscr{B}_0\,. \tag{2.1}$$

Using $W(F^*\times G) = W(\{\alpha(\alpha(F))\}^*\times G)$ we obtain from (2.1)

[a] Permanent address: Institute of Theoretical Physics, University of Wroclaw, Wroclaw, Poland.

 0022-2488/88/051231-05$02.50

$$|W(F^*\times G)|^2 \leqslant W(F^*\times\alpha(F))W(\alpha(G)^*\times G). \tag{2.2}$$

By condition (3) of W'3, $W(F^*\times\alpha(F)) = W(\alpha(F)^*\times F)$, so that, for $F,G\in\mathscr{B}_0$,

$$|W(F^*\times G)| \leqslant p_\alpha(F)p_\alpha(G). \tag{2.3}$$

Since $\ker p_\alpha = \{F\in\mathscr{B}:\ p_\alpha(F) = 0\}$, from the extension of (2.3) we have $\ker p_\alpha \subset \mathscr{N}_W$. Defining $\mathscr{N}^0_W = \{F\in\mathscr{B}_0: W(F^*\times G) = 0,\ \forall G\in\mathscr{B}\}$ we see that $\mathscr{N}_W = \overline{\mathscr{N}^0_W}$. Since $\mathscr{N}^0_W \subset \ker p_\alpha$ and p_α is $\mathscr{B}$ continuous, $\mathscr{N}_W = \overline{\mathscr{N}^0_W} \subset \ker p_\alpha$.

The extended seminorm p_α on $\mathscr{B}$ satisfies the characteristic assumptions of the framework discussed in Ref. 7. Thus all the results proved there follow. Actually, in this case we get more, namely the Hilbert space $\mathscr{K}^W = \overline{\mathscr{D}^W}$, obtained by closing $\mathscr{D}^W$ with respect to the Hilbert topology defined by p_α, is a *Krein space*, i.e., there is a metric operator η such that, for any $F,G\in\mathscr{B}_0$,

$$([F]_W,\eta[G]_W)_\alpha = \langle [F]_W,[G]_W\rangle = W(F^*\times G) \tag{2.4}$$

and $\eta^2 = 1$. (Here $[F]_W$ denotes the equivalence class of F with respect to the Wightman kernel $\mathscr{N}_W$.) Such Hilbert space structure will be briefly called a *Krein structure*. In fact we have the following theorem.

Theorem 2.2: The Hilbert space structure defined by p_α is a Krein space structure.

Proof: Since $\mathscr{B}_0$ is dense in $\mathscr{B}$, $\mathscr{D}^W_0 \equiv \mathscr{B}_0/\mathscr{N}_W$ is dense in $\mathscr{D}^W \equiv \mathscr{B}/\mathscr{N}_W$ with respect to the quotient topology of $\mathscr{B}$ modulo $\mathscr{N}_W$ on $\mathscr{D}^W_0$ we define a positive inner product

$$([F]_W,[G]_W)_\alpha \equiv W(\alpha(F)^*\times G). \tag{2.5}$$

Next we define

$$\eta[F]_W \equiv [\alpha(F)]_W, \quad \forall F\in\mathscr{B}_0. \tag{2.6}$$

Equation (2.6) is well defined since α maps equivalence classes into equivalence classes. Furthermore, $\eta^2[F]_W = [\alpha(\alpha(F))]_W = [F]_W$, i.e., $\eta^2 = 1$. Moreover

$$\langle \eta[F]_W,[G]_W\rangle = \langle [F]_W,\eta[G]_W\rangle$$

and

$$\langle [F]_W,[G]_W\rangle = ([F]_W,\eta[G]_W)_\alpha, \quad F,G\in\mathscr{B}_0. \tag{2.7}$$

Completing $\mathscr{D}^W_0$ with respect to the norm p_α we obtain the Hilbert space $\mathscr{K}^W$. Since $\mathscr{D}^W_0$ is p_α-dense in $\mathscr{D}^W$ (by $\mathscr{B}$-continuity), we have $\overline{\mathscr{D}^W_0} = \overline{\mathscr{D}^W} = \overline{\mathscr{K}^W}$, so that the local states $\mathscr{D}^W$ are dense in $\mathscr{K}^W$. The operator η defined $\mathscr{D}^W_0$ can be extended to $\mathscr{K}^W$ and for any $\Psi,\Phi\in\mathscr{K}^W$ we have

$$\langle\Psi,\Phi\rangle = (\Psi,\eta\Phi)_\alpha. \tag{2.8}$$

Hence $\mathscr{K}^W$ is a Krein space. Since, by (1), (3), $W(\alpha(G)) = W(G)$, we have

$$W((\alpha(1)-1)^*\times G) = W(\alpha(1)^*\times G) - W(G) = 0,$$

i.e., $[\alpha(1)]_W = [1]_W$, so that the vacuum vector $\Psi_0 = [1]_W$ is η-invariant, i.e., $\eta\Psi_0 = \Psi_0$. (In general, however, η does not leave $\mathscr{D}^W$ invariant.)

Remark: A stronger property then Krein positivity is the α positivity discussed in Ref. 8, namely when α is an automorphism of the whole Borchers algebra $\mathscr{B}$. This latter is the case of free QED, see Sec. IV, but not the case of interacting QED,[4] for which we believe that our more general Krein structure is relevant.

B. Euclidean case

Let $\{S_n\}$ be the set of Schwinger functions obtained by analytic continuation from the Wightman functions $\{W_n\}$. They satisfy the following axioms[12] (again we restrict ourselves to the case of a scalar Hermitian field).

OS1 *(temperedness):* $\forall_n$, $S_n\in\mathscr{S}_0(\mathbb{R}^{4n})'$ and obeys the Hermiticity property $\overline{S_n(f)} = S_n(\theta f^*)$.

OS2 *(Euclidean covariance):* $\forall n$, for each $R\in SO(4)$, $a\in\mathbb{R}^4$, $S_n(f_{\{a,R\}}) = S_n(f)$.

OS4 *(symmetry):* $\forall n$, $S_n(x_{\pi(1)},\dots,x_{\pi(n)}) = S_n(x_1,\dots,x_n)$ for all permutations of n elements.

The set $\{S_n\}$ of Schwinger functions is *Krein positive* if[13] it satisfies the following axiom.

OS3′ *(Krein positivity):* There exists a mapping α_s of the Borchers algebra $\mathscr{B}_+$ into itself such that $\forall F_+,G_+\in\mathscr{B}_+$

(1) $S(\{\theta\alpha_s(\alpha_s(F_+))\}^*\times G_+) = S(\{\theta F_+\}^*\times G_+)$;

(2) $S(\{\theta\alpha_s(F_+)\}^*\times F_+) \geqslant 0$;

(3) $S(\{\theta\alpha_s(F_+)\}^*\times G_+) = S(\{\theta F_+\}^*\times\alpha_s(G_+))$;

(4) $p_{\alpha_s}(F_+) \equiv S(\{\theta\alpha_s(F_+)\}^*\times F_+)^{1/2}$ is continuous with respect to the topology of $\mathscr{B}_+$ (briefly $\mathscr{B}_+$-continuous).[7]

Proposition 2.3: The seminorm p_{α_s} is nondegenerate, i.e.,

$$\ker p_{\alpha_s} = \mathscr{N}_s \equiv \{F_+\in\mathscr{B}_+:\ S(\{\theta F_+\}^*\times G_+) = 0,\ \forall G_+\in\mathscr{B}_+\}.$$

Proof: As in the relativistic case, by using conditions (2) and (3) of OS3′, we can show that

$$|S(\{\theta F_+\}^*\times G_+)| \leqslant p_{\alpha_s}(F_+)p_{\alpha_s}(G_+). \tag{2.9}$$

Hence $\ker p_{\alpha_s} \subset \mathscr{N}_s$. On the other hand, using (3) OS3′ we see that if $F_+\in\mathscr{N}_s$ then

$$S(\{\theta\alpha_s(F_+)\}^*\times F_+) = S(\{\theta F_+\}^*\times\alpha_s(F_+)) = 0,$$

so that $F_+\in\ker p_{\alpha_s}$, i.e., $\mathscr{N}_s \subset \ker p_{\alpha_s}$.

Remark: The seminorm p_{α_s} in a natural way defines a Hilbert space structure in the Euclidean vector space.[7] As in the relativistic case, the Hilbert space closure $\mathscr{K}^s$ of Euclidean local states $\mathscr{D}^s = \mathscr{B}_+/\mathscr{N}_s$ is a Krein space with the metric operator η_s. The construction is analogous to that of the relativistic case.

III. ANALYTIC CONTINUATION OF KREIN STRUCTURES

As discussed in Ref. 7, it is relevant to establish a connection between the Hilbert space structure in the relativistic case and that of the Euclidean case. In particular, it is relevant to know under which conditions the existence of a Krein structure in the relativistic case guarantees the existence of a Krein structure in the Euclidean case and vice versa. The solution of this problem is discussed in this section.

A. From relativistic to Euclidean Krein structure

If $\{W_n\}$ is the set of Wightman functions satisfying W1, W2, W4, and W5, and the Krein positivity W3′ for some α and $\mathscr{B}_0$, then the Wightman functions $\widetilde{W}_n^d$ defined on $\mathscr{S}(\overline{\mathbb{R}}_+^{4n})$ satisfy a Krein positivity condition with respect to the subalgebra $\widehat{\mathscr{B}}_0 \subset \widehat{\mathscr{B}}$, where the mapping $\hat{}$ is defined by

$$\hat{f}_n(q_1,...,q_n) = \tilde{f}_{nd}(q_1,...,q_n) \downharpoonright \{q_k^0 \geqslant 0\},$$

with $f_d(x_1, x_2 - x_1,..., x_n - x_{n-1}) \equiv f(x_1,...,x_n)$, and $\hat{\alpha}$ is defined by $\hat{\alpha}(\widehat{F}) \equiv \alpha(f)$. Since the Wightman functions and Schwinger functions are related by

$$S(\theta F_+^* \times G_+) = \widetilde{W}^d(\check{F}_+^* \times \check{G}_+), \tag{3.1}$$

where $\check{f}_{+n}(q_1,...,q_n) = f^{\mathrm{FL}}_{+nd}(q_1,...,q_n) \downharpoonright \{q_k^0 \geqslant 0\}$, and FL denotes the Fourier–Laplace transform, we have the following proposition.

Proposition 3.1: If the Wightman functions $\{W_n\}$ are Krein positive with respect to $\mathscr{B}_0 = \{F \in \mathscr{B} : \widehat{F} \in \check{\mathscr{B}}_+\}$, then the corresponding Schwinger functions are Krein positive.

Proof: By denoting by ϕ the inverse of $\check{}$ we define $\alpha_s(F_+) = \phi(\check{\alpha}(\check{F}_+))$ so that $\alpha_s : \mathscr{B}_+ \to \mathscr{B}_+$ and by (3.1)

$$S(\{\theta\alpha_s(F_+)\}^* \times F_+) = \widetilde{W}^d(\hat{\alpha}(\check{F}_+)^* \times \check{F}_+) \geqslant 0.$$

From the corresponding properties of $\widetilde{W}^d$ we also obtain conditions (1) and (3) of OS3′. Since p_α is $\mathscr{B}$ continuous and nondegenerate, $\hat{p}_{\hat{\alpha}}(\widehat{F}) = \widetilde{W}^d(\hat{\alpha}(\widehat{F})^* \times \widehat{F})^{1/2}$ is $\mathscr{B}(\overline{\mathbb{R}}_+^4)$ continuous,[7] so that

$$p_{\alpha_s}(F_+) = S(\{\theta\alpha_s(F_+)\}^* \times F_+)^{1/2} = \hat{p}_{\hat{\alpha}}(\check{F}_+)$$

is $\mathscr{B}(\overline{\mathbb{R}}_+^4)$ continuous. Then $\mathscr{B}_+$ continuity follows from the continuity of the mapping $F_+ \to \check{F}_+$.

B. From Euclidean to relativistic Krein structure

Proposition 3.2: If the Schwinger functions $\{S_n\}$ are Krein positive and the corresponding seminorm p_{α_s} is $\mathscr{B}(\overline{\mathbb{R}}_+^4)$ continuous, we can construct the Wightman functions satisfying the Krein positivity condition.

Proof: Since p_{α_s} is $\mathscr{B}(\overline{\mathbb{R}}_+^4)$ continuous, the existence of Wightman functions follows from the arguments of Ref. 7. We then define a mapping $\check{\alpha}_s$ on $\check{\mathscr{B}}_+$ by $\check{\alpha}_s(\check{F}_+) \equiv (\alpha_s(F_+))\check{}$, so that $\check{\alpha}_s : \check{\mathscr{B}}_+ \to \check{\mathscr{B}}_+$. Moreover, by (3.1) and OS3′(2),

$$\widetilde{W}^d(\hat{\alpha}_s(\check{F}_+)^* \times \check{F}_+) = S(\{\theta\alpha_s(F_+)\}^* \times F_+) \geqslant 0.$$

Hence $\widetilde{W}^d$ satisfies a Krein positivity condition with respect to the subalgebra $\check{\mathscr{B}}_+$ and with the mapping $\check{\alpha}_s$. Furthermore, since there is a natural identification $\mathscr{I}$ between $\check{\mathscr{B}}_+/\mathscr{N}_W^d$ and $\mathscr{B}_0/\mathscr{N}_W$, the mapping $\alpha\,\check{\alpha}_s$ defines a mapping of the equivalence classes of $\mathscr{B}_0$ by $\hat{\alpha} = \mathscr{I}\check{\alpha}_s\mathscr{I}^{-1}$. Such a mapping can be lifted from $B_0/\mathscr{N}_W$ to B_0 (in a nonunique way!)[14] and such a lifted map will be denoted by α. Then $\{W_n\}$ satisfy a Krein positivity condition for $\mathscr{B}_0$ and α. The density of $\mathscr{B}_0$ in $\mathscr{B}$ is standard.[15]

IV. FREE QED IN LOCAL GAUGES

A simple example, which realizes the structure discussed above, is provided by the Gupta–Bleuler formulation of free QED.

The theory is defined by the two-point function

$$W_2^{\mu\nu}(x_1,x_2) = -(2\pi)^{-3} g^{\mu\nu} \int d\Omega_0(k) e^{ik(x_1 - x_2)}, \tag{4.1}$$

with $d\Omega_0(k) = d\,\mathbf{k}/|\mathbf{k}|$. If we introduce the Borchers algebra $\mathscr{B}$ generated by vector test functions $f(x) = \{f^\mu(x)\}$, $\mu = 0, 1, 2, 3$, $f^\mu(x) \in \mathscr{S}(\mathbb{R}^4)$, then the Wightman functions $\{W_n\}$ are defined as follows:

$$W_0 = 1, \quad W_{2n+1} = 0,$$

$$W_2(f \times g) = \sum \int dx_1\, dx_2 f_\mu(x_1) g_\nu(x_2) W_2^{\mu\nu}(x_1,x_2),$$

$$W_{2n}(f_1 \times \cdots \times f_{2n}) = \sum_{(ij)} \prod_{\nu=1}^{n} W_2(f_{i_\nu} \times f_{j_\nu}), \tag{4.2}$$

where the sum is over all partitions of the indices $(1,...,2n)$ into distinct pairs $(i_1, j_1),...,(i_n, j_n)$ with $i_\nu < j_\nu$. It is easy to check that the Wightman functions defined above are Krein positive with respect to the entire Borchers algebra $\mathscr{B}$ with α defined by

$$\begin{aligned} &\alpha(f_1 \otimes \cdots \otimes f_n) = \eta_1 f_1 \otimes \cdots \otimes \eta_1 f_n, \\ &(\eta_1 f)_\mu = -g_\mu^\nu f_\nu . \end{aligned} \tag{4.3}$$

In fact, α defined by (4.3) is an automorphism of $\mathscr{B}$, $\alpha^2 = 1$ and

$$W_2(\alpha(f)^* \times f) = \sum \int d\Omega_0(k) |f_\mu(k)|^2 \geqslant 0.$$

Furthermore, $W_2(\alpha(f)^* \times g) = W_2(f^* \times \alpha(g))$. Finally $p_\alpha(f)$ is a $\mathscr{B}$-continuous Hilbert seminorm that is also $\mathscr{B}(\overline{\mathbb{R}}_+^4)$ continuous.

Remark 1: The mapping α, which defines the Krein positivity, is not unique; for example, we can put in (4.3) the matrix η_{1t}:

$$\eta_{1t} = \begin{pmatrix} -(\cosh^2 t + \sinh^2 t) & 2\cosh t \sinh t & 0 \\ -2\cosh t \sinh t & \cosh^2 t + \sinh^2 t & \\ 0 & & 1 \end{pmatrix},$$

for any $t \in \mathbb{R}$. In this way we obtain a one-parameter family of α's leading to nonequivalent Krein structures for $\{W_n\}$, i.e., the topologies defined by two α's of such a family are inequivalent. Such arbitrariness may be resolved by taking into account the properties of the physical subspace.

Remark 2: Since the mapping α defined by (4.3) leaves invariant the subalgebra $\mathscr{B}_0 = \{F \in \mathscr{B} : \widehat{F} \in \check{\mathscr{B}}_+\}$, $\{W_n\}$ is also Krein positive with respect to this subalgebra, and the theory developed in the previous section can be applied.

Now we pass to the discussion of the analytic continuation of this theory. The fact that we are dealing with a vector case requires some care and it seems worthwhile to spell out the basic steps. Let us denote by z^s the points in the extended permuted tube $\mathscr{T}^{\text{ext},p}$ of the form $z^s = (iz^0, z)$, z^0, $z^j \in \mathbb{R}$ (Schwinger points). Then we define the corresponding Euclidean points $x^E \equiv I^{-1} z^s = (z^0, z)$, with $I = \mathrm{diag}(i,1,1,1)$. In the scalar case, the Schwinger functions S_n are defined in the following way:

$$S_n(x_1^E,...,x_n^E) \equiv W_n(Ix_1^E,...,Ix_n^E), \tag{4.4}$$

where W_n is the Wightman function analytic in the tube $\mathscr{T}^{\text{ext},p}$. Since the W_n are invariant with respect to all complex

Lorentz transformations, the S_n are invariant with respect to the orthogonal group: if $\Lambda\in L_+(\mathbb{C})$ and $g\bar{\Lambda}=\Lambda g$ then $I^{-1}\Lambda I\in SO(4)$ and

$$S_n((I^{-1}\Lambda I)x_1^E,...,(I^{-1}\Lambda I)x_n^E) = W_n(\Lambda Ix_1^E,...,\Lambda Ix_n^E) = W_n(Ix_1^E,...,Ix_n^E) = S_n(x_1^E,...,x_n^E)\,.$$

In the vector case, the definition of the Schwinger functions,

$$S_{\mu_1,...,\mu_n}(x_1^E,...,x_n^E) = \sum (I^{-1})_{\mu_1\nu_1}\cdots(I^{-1})_{\mu_n\nu_n}W_{\nu_1,...,\nu_n}(Ix_1^E,...,Ix_n^E) \quad (4.5)$$

[which naturally generalizes Eq. (4.4)], guarantees the covariance of the Schwinger functions under SO (4):

$$S_{\mu_1,...,\mu_n}((I^{-1}\Lambda I)x_1^E,...,(I^{-1}\Lambda I)x_n^E) = \sum (I^{-1}\Lambda I)_{\mu_1\nu_1...}(I^{-1}\Lambda I)_{\mu_n\nu_n}S_{\nu_1...\nu_n}(x_1^E,...,x_n^E)$$

if $\{W_n\}$ is $L_+(\mathbb{C})$ covariant. In the case of free QED, we obtain

$$S_{\mu\nu}(x_1^E,x_2^E) = \frac{\delta_{\mu\nu}}{(2\pi)^3}\int \frac{d\mathbf{k}}{|\mathbf{k}|}e^{-|\mathbf{k}|(x_1^0-x_2^0)-i\mathbf{k}\cdot(\mathbf{x}_1-\mathbf{x}_2)}\,. \quad (4.6)$$

From the two-point Schwinger function (4.6) we construct the Schwinger functions $\{S_n\}$ similarly as in the relativistic case. Moreover, by analytic continuation we have

$$S^d_{\mu_1\mu_2}(\xi) = \sum \int e^{-(\xi^0q^0-i\boldsymbol{\xi}\cdot\mathbf{q})}(I^{-1})_{\mu_1\nu_1}(I^{-1})_{\mu_2\nu_2} \times W^d_{\nu_1\nu_2}(q)dq^4\,. \quad (4.7)$$

To discuss the relation between Wightman and Schwinger functions we must define the operator θ in such a way that (3.1) holds. In particular for the two-point functions we must have

$$S((\theta f)^*\times g) = \tilde{W}^d(\check{f}^*\times\check{g})\,,$$

for $f=\{f_\mu\}$, $g=\{g_\mu\}$, $f_\mu,g_\mu\in\mathscr{S}_+(\mathbb{R}^4)$, $\mu=0,1,2,3$. We look for the operator θ of the form

$$(\theta f)_\mu(x) = \sum M_{\mu\nu}f_\nu(rx)\,.$$

Using (3.1) and (4.7) we obtain that $M=I^2=\mathrm{diag}(-1,1,1,1)=M^{-1}$. For the generic n-point test function we clearly have

$$\theta f_n(x_1,...,x_n) = M\otimes\cdots\otimes Mf_n(rx_1,...,rx_n)\,.$$

Remark: The operator θ defined above realizes the natural representation of time reversal r in the space of tensor test functions.

As a consequence of (4.6) and the definition of θ we have the following proposition.

Proposition 4.1: The set of Schwinger functions $\{S_n\}$, corresponding to the Wightman functions for QED in the Gupta–Bleuler formulation, do not satisfy the OS-positivity condition.

Remark: Even if $\{S_n\}$ does not satisfy the OS-positivity condition, they are Nelson–Symanzik positive, i.e., $S(F^*\times F)\geqslant 0$.

Proposition 4.2: The set of Schwinger functions for free QED in the Gupta–Bleuler formulation satisfies the Krein positivity condition.

Proof: Define the mapping α_s of $\mathscr{B}_+$ by

$$\alpha_s(f) = Mf, \quad \text{for } f=\{f_\mu\},\quad f_\mu\in\mathscr{S}_+(\mathbb{R}^4),\quad \mu=0,1,2,3,$$

and by the tensor product of M in tensor test functions. By (4.6) we have

$$S_2(\{\theta\alpha_s(f)\}^*\times f) = \frac{1}{(2\pi)^3}\sum_{\mu=0}^{3}\int\frac{d\mathbf{k}}{|\mathbf{k}|}\left|\int dx_0 f_\mu(x_0,k)e^{-x_0|\mathbf{k}|}\right|^2 \geqslant 0\,.$$

Similarly

$$S_2(\{\theta\alpha_s(f)\}^*\times g) = S_2(\{\theta f\}^*\times\alpha_s(g))\,.$$

Moreover, $\alpha_s^2=1$ and $p_{\alpha_s}(f)=S_2(\{\theta\alpha_s(f)\}^*\times f)^{1/2}$ is obviously $\mathscr{B}_+$ continuous and $\mathscr{B}(\bar{\mathbb{R}}^4_+)$ continuous.

Remark 1: The mapping α_s defined above is the simplest one realizing the Krein positivity of $\{S_n\}$. As in the relativistic case, the Krein structure for a given set of Schwinger functions is not uniquely defined. Again the property of the physical states may be used to resolve such arbitrariness.

Remark 2: It is worthwhile to remark that the lack of OS positivity is not merely due to a wrong choice of the θ reflection operator, since the representation of θ is dictated by the representation of the time reversal operator which in turn is fixed by the analytic continuation of Wightman functions. As a matter of fact, setting $\theta'=\theta\alpha_s$ one has $S_2((\theta' f)^*\times f)\geqslant 0$; however, θ' does not provide the correct representation of the time reversal.

Remark 3: Since the mapping α commutes with the mappings $\hat{}$ and $\check{}$ (see Sec. III), one can show that α, $\hat{\alpha}$, and α_s are represented by the same matrix (each acting in the correponding space $\mathscr{B}$, $\hat{\mathscr{B}}$, and $\mathscr{B}_+$).

Remark 4: Since the two-point Schwinger function S_2 is Nelson–Symanzik positive (even if it is not OS positive), we can construct the Gaussian stochastic process $A(f)$, indexed by vector functions $f=\{f_\mu\}$, $f_\mu\in\mathscr{S}_{\mathbb{R}}(\mathbb{R})^4$, $\mu=0,1,2,3$, with covariance $E(A(f)A(g))\equiv S_2(f\times g)$.[16] So we have the probability space (Ω,Σ,μ) and $A(f)$ is a random variable on it. Furthermore, we can construct the representation $\mathscr{U}$ of the Euclidean group (without reflections) by means of unitary operators defined on $L_2(\Omega,\Sigma,\mu)$ and such that

$$\mathscr{U}(a,R)A(f)\mathscr{U}(a,R)^{-1} = A(f_{\{a,R\}})\,. \quad (4.8)$$

The Schwinger functions of the free QED can be constructed from the random variables $A(f)$ by

$$S_n(f_1\times\cdots\times f_n) = \int A(f_1)(\omega)\cdots A(f_n)(\omega)d\mu(\omega)\,.$$

However, $A(f)$ is not a Euclidean vector field in the sense of Nelson,[16] since the reflection property ensuring the OS positivity does not hold. More specifically, the representation of the "time" reflection does not lead to a reflection property that yields the OS positivity. As a consequence of this, the reconstruction of the Wightman functions cannot be done in the standard way. One can reconstruct the Wightman functions by following the strategy discussed in Sec. III, by using the Krein positivity. The result is that the Wightman function so obtained gives rise to a nonunitary representation of the Poincaré group (η-unitary representation), even if the Nelson formulation leads to a unitary representation of the Euclidean group [Eq. (4.8)].

[1]F. Strocchi and A. S. Wightman, J. Math. Phys. **15**, 2198 (1974).

[2]G. Morchio and F. Strocchi, Ann. Inst. H. Poincaré, Sect. A **33**, 251 (1980).

[3]F. Strocchi, Phys. Rev. D **17**, 2010 (1978); Commun. Math. Phys. 56, 57 (1977).

[4]G. Morchio and F. Strocchi, Nucl. Phys. B **211**, 471 (1983); **232**, 547 (1984); "Infrared problem, Higgs phenomena and long range interaction," *Fundamental Problems of Gauge Field Theory*, Erice, 1985, edited by G. Velo and A. S. Wightman (Plenum, New York, 1986).

[5]G. Morchio and F. Strocchi, Ann. Phys. (NY) **168**, 27 (1986); **172**, 267 (1986).

[6]G. Morchio, D. Pierotti, and F. Strocchi, "Infrared and vacuum structure in two dimensional local quantum field theory models. I, II," SISSA preprints 84/87/ EP and 105/87/ EP.

[7]L. Jakobczyk and F. Strocchi, "Euclidean formulation of quantum field theory without positivity," SISSA preprint, 59/87/FM.

[8]L. Jakobczyk, J. Math. Phys. **25**, 617 (1984).

[9]We shall denote the axiom which replaces Wightman positivity W3 by W3′, as in Ref. 7 even if the present form is stronger than the one discussed in Ref. 7.

[10]H. J. Borchers, "Algebraic aspects of Wightman field theory," in *Statistical Mechanics and Field Theory*, edited by R. N. Sen and C. Weil (Wiley, New York, 1972).

[11]Given $F \in \mathscr{B}$, we denote by f_n its n-variable component.

[12]K. Osterwalder and R. Schrader, Commun. Math. Phys. **31**, 83 (1973); **42**, 281 (1975). For general notation see Ref. 7.

[13]We shall denote the axiom which replaces the OS positivity OS3 by OS3′, as in Ref. 7, even if the present form is stronger than the one discussed in Ref. 7.

[14]But different α's define the same positive inner product and therefore the same metric operator η.

[15]B. Simon, *The $P(\phi)_2$ Euclidean (Quantum) Field Theory* (Princeton U.P., Princeton, NJ, 1974).

[16]E. Nelson, "Probability theory and Euclidean field theory," in *Constructive Quantum Field Theory*, Erice, 1973 edited by G. Velo and A. S. Wightman (Springer, Berlin, 1973); F. Guerra, "Local algebras in Euclidean quantum field theory," in *Symposia Matematics*, Vol. XX (Academic, London, 1976), and references therein.

Stochastic field theory and finite-temperature supersymmetry

Pradip Ghosh
Maharaja Manindra Chandra College, Calcutta-700002, India
Pratul Bandyopadhyay
Indian Statistical Institute, Calcutta-700035, India

(Received 5 January 1987; accepted for publication 18 November 1987)

The finite-temperature behavior of supersymmetry is considered from the viewpoint of stochastic field theory. To this end, it is considered that Nelson's stochastic mechanics may be generalized to the quantization of a Fermi field when the classical analog of such a field is taken to be a scalar nonlocal field where the internal space is anisotropic in nature such that when quantized this gives rise to two internal helicities corresponding to fermion and antifermion. Stochastic field theory at finite temperature is then formulated from stochastic mechanics which incorporates Brownian motion in the external space as well as in the internal space of a particle. It is shown that when the anisotropy of the internal space is suppressed so that the internal time ξ_0 vanishes and the internal space variables are integrated out one has supersymmetry at finite temperature. This result is true for $T = 0$, also. However, at this phase equilibrium will be destroyed. Thus for a random process van Hove's result involving quantum mechanical operators, i.e., that when supersymmetry remains unbroken at $T = 0$ it will also remain unbroken at $T \neq 0$, occurs. However, this formalism indicates that when at $T = 0$ broken supersymmetry results, supersymmetry may be restored at a critical temperature T_c.

I. INTRODUCTION

The finite-temperature behavior of supersymmetry has been treated with great importance by many authors.[1] Apparently we may think that supersymmetry is automatically broken at high temperature since bosons and fermions obey different statistics at high temperature. However van Hove[2] has pointed out that the observed "breakdown" of supersymmetry is only apparent since a careful definition of matrix elements for operators containing the anticommuting Grassmann parameters yields the opposite result.

All these arguments have been made using the proper definition of quantum mechanical operators. However, it is also known that supersymmetry may be associated with random phenomena and thus the stochastic quantization procedure involves hidden supersymmetry. Indeed, the stochastic quantization procedure of Parisi and Wu[3] introducing a fictitious time is found to have hidden supersymmetry. Again, in a recent paper[4] it has also been shown that Nelson's quantization procedure[5] involving universal Brownian motion also possesses hidden supersymmetry. To this end, Nelson's stochastic mechanics is generalized to the quantization of a fermion field when the classical analog of such a field is taken to be a scalar nonlocal field where the internal space is anisotropic in nature such that when quantized this gives rise to two internal helicities corresponding to fermion and antifermion. This also helps us to have a relativistic generalization of Brownian motion processes. This procedure gives the interesting result that when the internal variable is suppressed, supersymmetry arises. This indicates that when the Euclidean Markov field formalism is developed for a scalar particle from stochastic mechanics integrating out the internal space variables, we have hidden supersymmetry. This makes Nelson's stochastic quantization procedure equivalent to that of Parisi and Wu in the sense that both these precedures involve hidden supersymmetry.

However, these results have been derived at $T = 0$. We shall study here supersymmetric features of Nelson's stochastic quantization procedure at $T \neq 0$. Indeed, it will be shown that even at $T \neq 0$, these results hold. This indicates that the crucial result derived by van Hove regarding the unbroken nature of supersymmetry at $T \neq 0$, when it is unbroken at $T = 0$, involving quantum mechanical operators is also valid for random phenomena generating supersymmetry.

II. STOCHASTIC FIELD THEORY AT FINITE TEMPERATURE

It has been shown that when an internal variable ξ_μ is incorporated in addition to the space-time variable x_μ to a quantum oscillator, an internal helicity is generated that corresponds to the fermion number of the system.[4] So to quantize a fermion we can start with a classical oscillator with an internal variable such that when quantized, this internal variable will give rise to internal helicity. We now want to apply Nelson's stochastic process to this system at finite temperature $T \neq 0$. Nelson's stochastic quantization procedure is based on the assumption that the configuration variable $q(t)$ is promoted to a Markov process. The conditions of the process are (i) existence of universal Brownian motion, and (ii) validity of the Euler–Lagrange equation. Since we are also dealing with internal space, we assume the existence of Brownian motion both in external and internal space. Hence we denote the configuration variable as $Q(t,\xi_0)$, where ξ_0 is the fourth component of the internal four-vector ξ_μ. The variable $Q(t,\xi_0)$ is assumed to be a separable function given by

$$Q(t,\xi_0) = q(t)q(\xi_0) . \tag{1}$$

We assume that the process $Q(t,\xi_0)$ satisfies the stochastic differential equations,

$$dQ_i(t,\xi_0) = b_i(Q(t,\xi_0),t,\xi_0)dt + d\omega_i(t) \,,$$
$$dQ_i(t,\xi_0) = b_i'(Q(t,\xi_0),t,\xi_0)d\xi_0 + d\omega_i(\xi_0) \,, \tag{2}$$

where $b_i(Q(t,\xi_0),t,\xi_0)$ and $b_i'(Q(t,\xi_0),t,\xi_0)$ correspond to certain velocity fields in the external and internal space and the $d\omega_i$ exhibit independent Brownian motion. Since the process is a true Markov process, $d\omega_i(t)(d\omega_i(\xi_0))$ does not depend on $Q(s,s')$ for $s\leqslant t$ $(s'\leqslant\xi^0)$. The expectation values satisfied by them at $T\neq 0$ are[6]

$$\langle d\omega_i(t)\rangle_{T\neq 0} = 0\,, \quad \langle d\omega_i(\xi_0)\rangle_{T\neq 0} = 0\,;$$

$$\langle d\omega_i(t)d\omega_j(t')\rangle_{T\neq 0} = \frac{\delta_{ij}}{\beta m}\sum_{n=-\alpha}^{\alpha} e^{i\omega_n(t-t')}\,dt\,dt'\,, \tag{3}$$

$$\langle d\omega_i(\xi_0)d\omega_j(\xi_0')\rangle_{T\neq 0} = \frac{\delta_{ij}}{\beta\pi^0}\sum_{n=-\alpha}^{\alpha} e^{i\omega_n(\xi_0-\xi_0')}\,d\xi_0\,d\xi_0'\,, \tag{4}$$

with

$$\omega_n = 2\pi n/\beta\hbar\,.$$

It is easily seen that in the limit $\beta\to\alpha$, the finite-temperature behavior of $d\omega$ reduces to the case at $T=0$. It is noted that these correlation functions have been chosen in such a way that the KMS condition for equilibrium states is satisfied. However, for a pure stochastic interpretation we may consider Eqs. (4) as a postulate. In above equations $\hbar$ is Planck's constant divided by 2π and m and π^0 are suitable constants.

The description is asymmetrical in both "external" and "internal" time but we can also write

$$dQ_i(t,\xi_0) = b_i^*(Q_i(t,\xi_0),t,\xi_0)dt + d\omega_i^*(t)\,, \tag{5}$$

$$dQ_i(t,\xi_0) = b_i'^*(Q_i(t,\xi_0),t,\xi_0)d\xi_0 + d\omega_i^*(\xi_0)\,, \tag{6}$$

where now ω^* has the same properties as ω except that $d\omega_i^*(t)$ $[d\omega_j^*(\xi_0)]$ are independent of $Q(s,s')$ for $s\geqslant t(s'\geqslant\xi_0)$.

Now we can derive the following moments of the configuration variables as $\langle q_i(t)\rangle = \langle q_i(\xi_0)\rangle = 0$:

$$\langle q_i(t)q_j(t')\rangle = \frac{\delta_{ij}}{\beta m}\sum_{n=-\alpha}^{\alpha}\frac{e^{i\omega_n(t-t')}}{\omega^2+\omega_n^2}\,,$$
$$\langle q_i(\xi_0)q_j(\xi_0')\rangle = \frac{\delta_{ij}}{\beta\pi^0}\sum_{n=-\alpha}^{\alpha}\frac{e^{i\omega_n(\xi_0-\xi_0')}}{\omega'^2+\omega_n^2}\,. \tag{7}$$

Thus the moment of product variable satisfy the following moments at $T\neq 0$,

$$\langle Q(t,\xi_0)\rangle = 0\,, \tag{8}$$

$$\langle Q_i(t,\xi_0)Q_j(t',\xi_0')\rangle = \frac{\delta_{ij}}{\beta m}\frac{1}{\beta\pi^0}\sum_{n=-\alpha}^{\alpha}\frac{e^{i\omega_n(t-t')}}{\omega^2+\omega_n^2}\sum_{n=-\alpha}^{\alpha}\frac{e^{i\omega_n(\xi_0-\xi_0')}}{\omega'^2+\omega_n^2}\,. \tag{9}$$

Now we want to get a relativistic version of stochastic field theory, i.e., we have to define a real stochastic scalar field $\Phi(x,t,\xi)$. Let us consider a bounded and smooth region G in R^3. Let $e_i(\mathbf{x})$ be the characteristic orthonormal set of eigenfunctions of the three-dimensional Laplacian in G, i.e.,

$$\Delta e_i(\mathbf{x}) = -k_i^2 e_i(\mathbf{x})\,. \tag{10}$$

A similar characteristic function $e_j(\boldsymbol{\xi})$ is defined for internal three-dimensional space so that

$$\Delta' e_j(\boldsymbol{\xi}) = -\pi_j^2 e_j(\boldsymbol{\xi})\,,$$

where

$$\Delta' = \frac{\partial^2}{\partial\xi_1^2}+\frac{\partial^2}{\partial\xi_2^2}+\frac{\partial^2}{\partial\xi_3^2}\,.$$

Then the stochastic nonlocal field $\Phi(x,t,\xi)$ can be generated by associating a stochastic oscillator with each $Q_i(t,\xi_0)$ and taking the limit $G\to R^3$, i.e.,

$$\Phi(x,t,\xi) = \sum_{i,j} q_i(t)e_i(\mathbf{x})q_j(\xi_0)e_j(\boldsymbol{\xi})\,. \tag{11}$$

The moments of $\Phi(x,t,\xi)$ are derived from the moments of $q(t), q(\xi_0)$ at $T\neq 0$ as follows:

$$\langle\Phi(x,t,\xi)\rangle = 0\,, \tag{12}$$

$$\langle\Phi(x,t,\xi)\Phi(x',t',\xi')\rangle = \frac{1}{(2\pi)^3}\sum_{n=-\alpha}^{\alpha}\int d^3\mathbf{k}\,e^{i\mathbf{k}(\mathbf{x}-\mathbf{x}')}\frac{1}{\beta m}\frac{e^{i\omega_n(t-t')}}{\omega^2+\omega_n^2}\times\frac{1}{(2\pi)^3}\sum_{n=-\alpha}^{\alpha}\int d^3\boldsymbol{\pi}\,e^{i\pi(\boldsymbol{\xi}-\boldsymbol{\xi}')}\frac{1}{\beta\pi^0}\frac{e^{i\omega_n(\xi_0-\xi_0')}}{\omega'^2+\omega_n^2}\,. \tag{13}$$

It is noted that in the limit $\xi_0=\xi_0'=0$ the correlation function (7) just reduces to

$$\langle q_i(\xi_0)q_j(\xi_0')\rangle = \frac{\delta_{ij}}{\beta\pi^0}\sum_{n=-\alpha}^{\alpha}\frac{1}{\omega'^2+\omega_n^2} = \frac{\delta_{ij}}{\beta\pi^0}\frac{\pi}{\omega'}\coth\left(\frac{1}{2}\beta\hbar\omega'\right), \tag{14}$$

where the relation used is

$$\sum_{n=-\alpha}^{\alpha}\frac{1}{n^2+a^2} = \frac{\pi}{a}(\coth\pi a)\,. \tag{15}$$

This is a function of temperature and can be taken outside second integral of Eq. (13) and integrating the rest over ξ space we get the correlation function as

$$\langle\Phi(x,t)\Phi(x',t')\rangle = \frac{1}{(2\pi)^3}\frac{1}{\beta\pi^0}\frac{\pi}{\omega'}\coth\left(\frac{1}{2}\beta\hbar\omega'\right)\frac{1}{(2\pi)^3}\times\sum_{n=-\alpha}^{\alpha}\int d^3\mathbf{k}\,e^{i\mathbf{k}(\mathbf{x}-\mathbf{x}')}\frac{1}{\beta m}\frac{e^{i\omega_n(t-t')}}{\omega^2+\omega_n^2}\,. \tag{16}$$

Now we want to show that when the anisotropic feature of the internal space-time corresponding to the variable ξ_μ is taken into account implicity we can obtain the fermionic propagator in Euclidean space-time. For this purpose the anisotropy is generated by two opposite orientations of internal variable ξ_μ (and hence of $\pi_\mu = i\,\partial/\partial\xi_\mu$) and consider that the two opposite internal helicities correspond to particle and antiparticle states.

Equations (12) and (13) are effectively a correlation function in eight-dimensional space-time, four-dimensional in the external space-time variable and four-dimensional in internal space-time. Introducing a mapping between external and internal space Eqs. (12) and (13) can be reduced to an effective four-dimensional expression in an external space-time variable so that the anisotropic feature of internal space is focused properly. The mapping is as follows:

$$k^2=(k',\pi);\quad x^2=(x',\xi)\quad \text{and}\quad m^2=m'\pi^0. \tag{17}$$

Here (A,B) denotes an Euclidean product and each component of $k(x)$ is

$$k_i=\sqrt{k'_i\pi_i}\,,\quad x_i=\sqrt{x'_i\xi_i}\,. \tag{18}$$

The correlation function of the new field variables can now be obtained from expressions (12) and (13) for a particular mode $n=1$:

$$\begin{aligned}\langle\bar\Phi(x,t,\xi)\bar\Phi(x',t',\xi')\rangle &= f(\beta)\frac{1}{(2\pi)^4}\int\frac{e^{i(\sqrt{k'\pi},x(\xi)-x'(\xi'))}}{(k',\pi)+(m'\pi^0)}d^4\sqrt{k'\pi}\,2\pi\delta(\omega_1-\sqrt{k'_0\pi_0})\\ &= f(\beta)\frac{1}{(2\pi)^4}\int\frac{d^4\sqrt{k'\pi}\,e^{i(\sqrt{k'\pi},x(\xi)-x'(\xi'))}\cdot 2\pi\delta(\omega_1-\sqrt{k'_0\pi_0})}{(i\sqrt{(k',\pi)}+\sqrt{m'\pi^0})(-i\sqrt{(k',\pi)}+\sqrt{m'\pi^0})}.\end{aligned}\tag{19}$$

This mapping shows that the behavior of the particle in the external space is a manifestation of the behavior of the internal constituents and the motion of the particle is governed by the motion of constituents in the internal space as a whole. According to our assumption anisotropy of the internal space is due to a direction vector fixed in internal space and this gives rise to internal helicity. The two opposite helicities can be taken to be represented by $i\sqrt{\pi}$ and $-i\sqrt{\pi}$ and correspond to particle and antiparticle states. So for a single helicity state depicting particle (or antiparticle) we should take $-i\sqrt{\pi}$ (or $i\sqrt{\pi}$) as a vanishing term. Taking $-i\sqrt{\pi}=0$ we see that the expression (18) reduces to the form

$$\langle\bar\Phi(x,t,\xi)\bar\Phi(x',t',\xi')\rangle=f(\beta)\frac{1}{(2\pi)^4}\int\frac{e^{i(k,(x-x'))}\,d^4k}{i\sqrt{k^2}+m}2\pi\delta(\omega_1-k_0)\,, \tag{20}$$

where we have taken $m=\pi^0=1$.

Now we can choose a matrix $(\gamma_\mu k_\mu+m)=(\not k+m)$ with two degenerate eigenvalues $\pm i\sqrt{k^2}+m$ that can be diagonalized by a unitary matrix U:

$$(\not k+m)=V^{-1}\begin{pmatrix} i\sqrt{k^2}+m & 0 & 0 & 0\\ 0 & i\sqrt{k^2}+m & 0 & 0\\ 0 & 0 & -i\sqrt{k^2}+m & 0\\ 0 & 0 & 0 & -i\sqrt{k^2}+m\end{pmatrix}U. \tag{21}$$

Thus we just get the fermionic propagator in Euclidean space-time,

$$\langle\bar\Phi(x,t,\xi)\bar\Phi(x,t,\xi)\rangle = f(\beta)\frac{1}{(2\pi)^4}\int\frac{e^{i(k,(x-x'))}\,d^4k}{\gamma_\mu k_\mu+m}2\pi\delta(\omega_1-k_0)\,. \tag{22}$$

Thus Eq. (22) gives the fermionic propagator in Euclidean space-time and the new field $\bar\Phi(x,t,\xi)$, where the anisotropic feature of internal space is manifested by internal helicity depicting a fermionic field.

III. SUPERSYMMETRY AT FINITE TEMPERATURE

In this section our aim is to establish the fact that supersymmetry is there at finite temperature when we apply the stochastic quantization process to a scalar nonlocal field when the anisotropic feature of the internal space vanishes. In view of this, we rewrite the stochastic differential equations involving the configuration variable $Q(t,\xi_0)$ in the form

$$dQ(t,\xi_0)=b(Q(t,\xi_0),t,\xi_0)dt+d\omega(t)\,, \tag{23}$$

$$dQ(t,\xi_0)=b'(Q(t,\xi_0),t,\xi_0)d\xi_0+d\omega(\xi_0)\,. \tag{24}$$

Here $d\omega(t)(d\omega(\xi_0))$ does not depend on $Q(s,s')$ for $s\leqslant t$ $(s'\leqslant\xi_0)$. Since the description is asymmetrical in both external and internal time we can also write

$$dQ(t,\xi_0)=b^*(Q(t,\xi_0),t,\xi_0)dt+d\omega^*(t)\,, \tag{25}$$

$$dQ(t,\xi_0)=b'^*(Q(t,\xi_0),t,\xi_0)d\xi_0+d\omega^*(\xi_0)\,, \tag{26}$$

where now ω^* has the same properties as ω except that $d\omega^*(t)(d\omega^*(\xi_0))$ are independent of $Q(s,s')$ for $s\geqslant t$ $(s'\geqslant\xi_0)$. From this we can define external current and osmotic velocity as

$$V(x,t,\xi)=\tfrac{1}{2}(b(x,t,\xi)+b^*(x,t,\xi)) \tag{27}$$

and

$$U(x,t,\xi)=\tfrac{1}{2}(b(x,t,\xi)-b^*(x,t,\xi))\,. \tag{28}$$

Similarly the internal current and "internal" osmotic velocities are defined as

$$V'=\tfrac{1}{2}\{b'(x,t,\xi)+b'^*(x,t,\xi)\} \tag{29}$$

and

$$U'=\tfrac{1}{2}\{b'(x,t,\xi)-b'^*(x,t,\xi)\}\,. \tag{30}$$

When the stochastic field $\Phi(x,t,\xi)$ is constructed from the random oscillators according to Eq. (11) and the external current velocity and the internal current velocity are expressed in the differential form

$$V=\frac{\delta W}{\delta\Phi}\,, \tag{31}$$

$$V' = \frac{\delta W'}{\delta \Phi}, \tag{32}$$

it is evident [from Eqs. (23), (25) and (24), (26)] that $\Phi(x,t,\xi)$ can be taken to satisfy the Langevin equation in external and internal space-time,

$$\frac{\partial \Phi(x,t,\xi)}{\partial t} = \frac{\delta W}{\delta \Phi} + \eta(x,t,\xi), \tag{33}$$

$$\frac{\partial \Phi(x,t,\xi)}{\partial \xi_0} = \frac{\delta W'}{\delta \Phi} + \eta'(x,t,\xi), \tag{34}$$

where $\eta(x,t,\xi)$ and $\eta'(x,t,\xi)$ are white noises. The relations satisfied by them at $T \neq 0$ are

$$\langle \eta(x,t,\xi) \rangle = \langle \eta'(x,t,\xi) \rangle = 0, \tag{35}$$

$$\begin{aligned} &\langle \eta(x,t,\xi)\eta(x',t',\xi') \rangle \\ &= \frac{1}{\beta m} \delta^3(\mathbf{x}-\mathbf{x}')\delta^3(\boldsymbol{\xi}-\boldsymbol{\xi}') \sum_{n=-\alpha}^{\alpha} e^{i\omega_n(t-t')} \\ &\times \sum_{n=-\alpha}^{\alpha} e^{i\omega_n(\xi_0-\xi_0')}, \end{aligned} \tag{36}$$

$$\begin{aligned} &\langle \eta'(x,t,\xi)\eta'(x',t',\xi') \rangle \\ &= \frac{1}{\beta \pi^0} \delta^3(\mathbf{x}-\mathbf{x}')\delta^3(\boldsymbol{\xi}-\boldsymbol{\xi}') \sum_{n=-\alpha}^{\alpha} e^{i\omega_n(t-t')} \\ &\times \sum_{n=-\alpha}^{\alpha} e^{i\omega_n(\xi_0-\xi_0')}. \end{aligned} \tag{37}$$

In the absence of internal anisotropy $\xi_0 = 0$, Eq. (34) reduces to

$$-\frac{\delta W'}{\delta \Phi} = \eta'(x,t). \tag{38}$$

Considering the equation in one dimension and writing U and W' we get

$$-\frac{\delta U}{\delta \Phi(x)} = \eta'(x). \tag{39}$$

Now for one particular mode $n = 1$, we can write, from Eq. (37) when $\xi_0 = \xi_0' = 0$ and the internal space variable $\boldsymbol{\xi}$ is integrated out,

$$\begin{aligned} \langle \eta'(x,t)\eta'(x',t') \rangle &= \frac{1}{\beta \pi^0} \delta^3(\mathbf{x}-\mathbf{x}') e^{i\omega_1(t-t')} \\ &= \frac{1}{\beta \pi^0} \delta^4(x-x') 2\pi \delta(\omega_1 - k_0). \end{aligned} \tag{40}$$

In the one-dimensional case, $\eta'(x)$ in Eq. (39) satisfies the relation

$$\langle \eta'(x)\eta'(x') \rangle = f(\beta)\delta(x-x'), \tag{41}$$

which is Gaussian in nature.

Now, as Parisi and Sourlas[7] have shown, if Eq. (39) has one and only one solution $\Phi_\eta(x)$, the expectation value of any function of $\Phi(x)$ is given by

$$\overline{F(\Phi)} = \int D\eta \exp\left(-\frac{1}{2}\int \eta^2(x)dx\right) F[\Phi_\eta(x)],$$

which can be written in the form

$$\begin{aligned} \overline{F(\Phi)} = &\int D\eta\, D\Phi \exp\left(-\frac{1}{2}\int \eta^2(x)dx\right) \\ &\times [F(\Phi(x)]\delta\left(\frac{\delta U}{\delta \Phi(x)} + \eta(x)\right)\det[V_{x,y}], \end{aligned} \tag{42}$$

where

$$U_{x,y} = \frac{\delta^2 U}{\delta \Phi(x) \delta \Phi(y)}. \tag{43}$$

Now writing $\det[U_{x,y}]$ as an integral over anticommuting variables, we have

$$\begin{aligned} \overline{F[\Phi\eta(x)]} = &\int D\Phi\, D\Psi\, F[\Phi(x)]\exp \\ &- \left[\frac{1}{2}\int dx\, U_x^2 + \int dx\, dy\, \bar{\Psi}(x) U_{x,y} \Psi(y)\right] \\ = &\int D\Phi\, D\Psi\, e^{-S(\Phi,\Psi)} F[\Phi], \end{aligned} \tag{44}$$

where

$$S(\Phi,\Psi) = \frac{1}{2}\int dx\, U_x^2 + \int dx\, dy\, \bar{\Psi}(x) U_{x,y} \Psi(y).$$

This action is invariant under the supersymmetry transformations

$$\begin{aligned} &\delta\Phi(x) = \bar{\epsilon}\Psi(x) + \bar{\Psi}(x)\epsilon, \\ &\delta\bar{\Psi}(x) = -\bar{\epsilon}U_x, \quad \delta\Psi(x) = -\epsilon U_x. \end{aligned} \tag{45}$$

This suggests that supersymmetry invariance also becomes implicit in Nelson's stochastic mechanics at finite temperature. This supports the results of van Hove involving quantum mechanical operators that when supersymmetry is unbroken at $T = 0$, it remains unbroken at $T \neq 0$. Our analysis suggests that this is true for supersymmetry generated from random phenomena, too.

IV. DISCUSSION

We have studied here stochastic field theories at finite temperature and the hidden supersymmetry found to be still there at $T \neq 0$ as it is at $T = 0$. The correlation functions of stochastic fluctuations have been chosen in such a way that the KMS condition is valid. That means that the equilibrium condition is built into the physical system considered here. But we must point out that, for the restoration of supersymmetry, we must assume that the anisotropy of the internal space is suppressed so that the internal time ξ_0 vanishes and the internal space variables can be integrated out; thus we get a local stochastic field only in terms of the external space-time variable. However, this destroys the equilibrium state and leads to a nonequilibrium condition. Indeed this becomes apparent from the fact that finite-temperature field theory in Minkowski space requires the existence of an extra field ("ghost field") that maintains time-reversal invariance (Z_2 symmetry) with the physical field as shown by Niemi and Semenoff.[8] When this Z_2 symmetry is broken we achieve a nonequilibrium state. In a recent paper we have shown that this ghost field can be associated with the field in the internal space when the stochastic nonlocal field is written as a thermal doublet $\binom{\Phi(x)}{\Phi(\xi)}$, where $\Phi(x)$ corresponds to the field in the external space and $\Phi(\xi)$ corresponds to the field in the internal space. Thus, when the internal time variable is taken to be vanishing and the internal space variables are integrated out, Z_2 symmetry is destroyed and we attain a nonequilibrium state. That means that equilibrium is destroyed in the supersymmetric phase.

It may be remarked here that for equilibrium states, the KMS condition not only in the external space but also in the internal space is an essential feature.[9] This is manifested in the fact that a stochastic field involving only the external space faces serious trouble since the two-point correlation function at $T=0$, $\langle\Phi(x,t)\Phi(x',t')\rangle$, involving only space-time variables, is not Lorentz invariant (rotationally invariant) when it is derived from the finite-temperature correlation function taking the limit $\beta\rightarrow\alpha$ (see Ref. 6). Indeed, the stochastic fluctuations operating at $T\neq0$ still have a residual effect at $T=0$ through the moment of the component oscillators. However, when the moments of the stochastic fields are determined incorporating two fields (one in the external variable and the other in the internal variables), this Lorentz invariance may be restored through CPT invariance as the symmetry manifested in this two-field formalism implies time reversal invariance, which again becomes equivalent to CP symmetry.

Finally we may point out that when the equilibrium is destroyed, the isotropic or the anisotropic feature of the internal space gets changed leading to the change in statistics of the thermal doublet and as such Z_2 symmetry is violated. This indicates that there may exist a critical temperature T_c below which supersymmetry remains broken, and, at $T>T_c$, supersymmetry is restored. This means that even when at $T=0$ supersymmetry remains broken implying an equilibrium state, we can achieve a critical temperature T_c when equilibrium is destroyed and supersymmetry is restored. This is a very significant result that we can achieve when supersymmetry is associated with random phenomena. Thus for a random process van Hove's result, i.e., when supersymmetry remains unbroken at $T=0$, it will also remain unbroken at $T\neq0$, is found to be valid implying that unbroken supersymmetry indicates a nonequilibrium state. Thus at $T\neq0$ this nonequilibrium condition will persist. But when at $T=0$ we have an equilibrium state implying broken supersymmetry a nonequilibrium state may be attained at a critical temperature when supersymmetry is restored. This spontaneous breakdown of supersymmetry at a critical temperature will then give rise to thermal doublets of opposite statistics, which will appear as new zero-energy modes as suggested by Matsumoto *et al.*[10] in the context of thermofield dynamics.

[1] L. Girardello, M. Grisaru, and P. Salomonson, Nucl. Phys. B **178**, 331 (1981).
[2] L. van Hove, Nucl. Phys. B **207**, 15 (1982).
[3] G. Parisi and Y. S. Wu, Sci. Sinica **24**, 483 (1981).
[4] P. Ghosh and P. Bandyopadhyay, Ann. Phys. (NY) **179**, 1 (1987).
[5] E. Nelson, Phys. Rev. **150**, 1079 (1966); *Dynamical Theory of Brownian Motion* (Princeton U. P., Princeton, NJ, 1967).
[6] S. M. Moore, J. Math. Phys. **21**, 2102 (1980).
[7] G. Parisi and N. Sourlas, Phys. Rev. Lett. **43**, 744 (1979).
[8] A.J. Niemi and G. W. Semenoff, Ann. Phys. (NY) **152**, 105 (1983).
[9] P. Bandyopadhyay, B. Bhattcharya, and P. Ghosh, submitted for publication.
[10] H. Matsumoto, N. Nakahara, Y. Nakano, and H. Umezawa, Phys. Lett. B **140**, 53 (1984).

The Virasoro algebra and group in four dimensions

Yao-Zhong Zhang
Center of Theoretical Physics, CCAST (World Laboratory) and Institute of Modern Physics, Northwest University, Xian, People's Republic of China

(Received 20 November 1987; accepted for publication 20 January 1988)

The Virasoro algebra and group are examined in $(3+1)$-dimensional theory with an Abelian gauge field coupled with the gravitational field. The two-cocycle for the group is constructed by using the so-called descent equation. Its infinitesimal version gives the Schwinger–Jackiw–Johnson term in the Virasoro algebra in $3+1$ dimensions. It appears that in the case of an external monopole field the Virasoro algebra and group in $3+1$ dimensions are the direct generalization of the standard ones in $1+1$ dimensions, respectively.

I. INTRODUCTION

The mixed anomaly in $3+1$ dimensions arises in the theory with matter fields coupled to both gravity and Abelian gauge fields.[1–3] If we stipulate that there is no gauge anomaly, then there is a gravitational anomaly also called the Einstein anomaly in the effective action because of the noninvariance of the space-time diffeomorphism. On the other hand, the commutator anomaly (Schwinger–Jackiw–Johnson term) of the Lie algebra of diffeomorphisms appears in $(1+1)$-dimensional theory with matter fields coupled to only gravity, and the algebra of infinitesimal diffeomorphisms has the form of a Virasoro algebra.[4–6] The group corresponding to this Virasoro algebra (called the Virasoro group) is also discussed by Jackiw and co-workers[7–10] and the present author,[11] who compute the two-cocycle for the group in different ways.

The purpose of this paper is to describe the Virasoro algebra and group in $(3+1)$-dimensional theory with a $U(1)$ gauge field coupled to the gravity field. The main point is the following. According to the family index theorem, the relevant term for a mixed anomaly can easily be determined. By using the descent equation one obtains the Schwinger–Jackiw–Johnson term in three space dimensions. Of particular interest is the case when the Abelian gauge field is fixed to be the monopole field. Then, as we shall see, the space fractionalization to $S^2\times S^1$ is obtained in a natural way and the algebra of infinitesimal diffeomorphisms will be exactly a Virasoro algebra in the radial variable; the angular coordinates appear in a trivial way as extra labels. The corresponding Virasoro group is also constructed and seen to be a direct generalization of the standard one in $1+1$ dimensions.

II. THEORY

I shall first give an introduction to the Virasoro algebra and group in $1+1$ dimensions, since the higher-dimensional case can be reduced to the former by factoring out the angular polar coordinates.

Let $\mathrm{Diff}(S^1)$ be the diffeomorphism group of smooth one-to-one maps $S^1\to S^1$, and $\mathrm{Lie}(\mathrm{Diff}(S^1))\equiv\mathrm{Vect}(S^1)$ its Lie algebra. Consider the Virasoro algebra $\mathrm{Vect}(S^1)^\wedge=\mathrm{Vect}(S^1)\oplus i\,\mathbb{R}$ with the commutator[11]

$$\left[\zeta\frac{d}{dx},\eta\frac{d}{dx}\right]=(\zeta'\eta-\eta'\zeta)\frac{d}{dx}+ic\beta(\zeta,\eta),\tag{1}$$

where $\zeta\, d/dx, \eta\, d/dx\in\mathrm{Vect}(S^1)$ and c is a real constant; $i\,\mathbb{R}$ commutes with everything. If $L_m=ie^{imx}\,d/dx$, then

$$[L_m,L_n]=(m-n)L_{m+n}+ic\beta(m,n).$$

At the group level we have a group $\mathrm{Ext}(\mathrm{Diff}(S^1))\equiv\mathrm{Diff}(S^1)^\wedge$, which has $\mathrm{Vect}(S^1)^\wedge$ as its Lie algebra. The multiplication law in $\mathrm{Diff}(S^1)^\wedge$ is

$$(G,\lambda)(H,\mu)=(G\circ H,\lambda\mu e^{i2\pi\alpha^2(G,H)}),\tag{2}$$

where $G,H\in\mathrm{Diff}(S^1)$, $\lambda,\mu\in S^1$ (circle group), and $\circ$ denotes composition; $\alpha^2(G,H)$ is the real-valued function of G and H. The infinitesimal version of the α^2 gives the central extension β. The associativity of the group multiplication law (2) implies that α^2 obeys the two-cocycle condition

$$(\Delta\alpha^2)(G,H,L)\equiv\alpha^2(H,L)-\alpha^2(G\circ H,L)+\alpha^2(G,H\circ L)-\alpha^2(G,H)=0 \mod n,\tag{3}$$

where Δ is the coboundary operator.[12]

The solution to (3) can be obtained as follows. Let $\Gamma^\rho_{\sigma\mu}$ be the Christoffel symbol. Define the Christoffel connection one-form $\Gamma=(\Gamma^\rho_\sigma)=(\Gamma^\rho_{\sigma\mu})dx^\mu$. The corresponding curvature two-form is $R=d\Gamma+\Gamma^2$. A finite diffeomorphism on S^n is simply the coordinate transformation[3]

$$\Gamma(x)\to\Gamma'(x')\equiv\Gamma^G=g(x)^{-1}(\Gamma+d)g(x),\tag{4}$$

where

$$(g(x)^{-1})^\alpha{}_\beta=\frac{\partial x'^\alpha}{\partial x^\beta},\quad x'=G(x).$$

Consider the diffeomorphism group $\mathrm{Diff}(S^n)$ of smooth one-to-one maps $S^n\to S^n$. The group multiplication is defined by the composition

$$GH\equiv G\circ H,\quad G\circ H=G(H(x)).\tag{5}$$

Let i $(=0,1,2,\ldots)$ specify some particular gauge, for example,

$$0\equiv\Gamma,\quad 1\equiv\Gamma^G=g(x)^{-1}(\Gamma+d)g(x),$$
$$2\equiv\Gamma^{GH}=(g(x)h(G(x)))^{-1}(\Gamma+d)(g(x)h(G(x))),\ldots .\tag{6}$$

The starting point is the Chern–Pontryagin form in four dimensions

$$\Omega_4^{-1}(R)=c\,\mathrm{tr}\,R^2,\tag{7}$$

where c is a normalization constant so that

 0022-2488/88/051241-03$02.50

$$\int_{S^4} \Omega_4^{-1}(R) = c\int_{S^4} \mathrm{tr}\, R^2 = \mathbb{Z}. \quad (8)$$

As is well-known, Ω_4^{-1} is exact,

$$\Omega_4^{-1} = d\Omega_3^0(\Gamma) \equiv d\Omega_3^0(0), \quad (9)$$

with Ω_3^0 being the Chern–Simons form in three dimensions:

$$\Omega_3^0(0) = c\,\mathrm{tr}(\Gamma\, d\Gamma + \tfrac{2}{3}\Gamma^3). \quad (10)$$

Applying the coboundary operator to (10) gives

$$(\Delta\Omega_3^0)(0,1) \equiv \Omega_3^0(1) - \Omega_3^0(0)$$
$$= cd\,\mathrm{tr}(\Gamma\, dg\, g^{-1}) - c\tfrac{1}{3}\mathrm{tr}(dg\, g^{-1})^3. \quad (11)$$

The three-form $C^{(3)} = -\tfrac{1}{3}\mathrm{tr}(dg\,g^{-1})^3$ is closed and locally $C^{(3)} = dH^{(2)}$ for some two-form H^2. If $g = \exp(u)$, then H^2 can be computed by[13]

$$H^{(2)}(g) = -\frac{1}{3}\int_0^t \mathrm{tr}(dg(x,t)g(x,t)^{-1})^3, \quad (12)$$

where $g(t,x) = \exp(tu(x))$. Thus we obtain the two-dimensional Einstein (or gravitational) anomaly

$$\Omega_2^1(0,1) = c\,\mathrm{tr}(\Gamma\, dg\, g^{-1}) + cH^{(2)}(g). \quad (13)$$

Similarly,

$$(\Delta\Omega_2^1)(0,1,2)$$
$$\equiv \Omega_2^1(1,2) - \Omega_2^1(0,2) + \Omega_2^1(0,1)$$
$$= c\,\mathrm{tr}(h(G(x))^{-1}\, dh(G(x)) dg(x) g(x)^{-1})$$
$$+ cH^{(2)}(g(x)) + cH^{(2)}(h(x))$$
$$+ cH^{(2)}(g(x)h(G(x))). \quad (14)$$

Let D^2 be some two-disk in four dimensions. Then any smooth one-to-one map G: $D^2 \to D^2$ can be restricted to the one G: $\partial D^2 \to \partial D^2$, where ∂D^2 is the boundary of D^2. We shall consider the case $S^1 = \partial D^2$. Let G and H be two diffeomorphisms of D^2 such that $G \circ H$: $D^2 \to D^2$; then we can define

$$\alpha^2(G,H) = \int_{D^2} (\Delta\Omega_2^1)(0,1,2)$$
$$= c\int_{D^2} \{\mathrm{tr}(h(G(x))^{-1}\, dh(G(x)) dg(x) g(x)^{-1})$$
$$+ H^{(2)}(g(x)) + H^{(2)}(h(x))$$
$$+ H^{(2)}(g(x)h(G(x)))\}. \quad (15)$$

The analogy with Refs. 14 and 15 shows that modulo an integer Z the function α^2 is independent of the chosen disk D^2. The two-cocycle condition can be checked similarly.

In the infinitesimal vicinity of the identity $g(x) = 1 + u(x)$, we obtain

$$\alpha^2(G,H) = c\int_{S^1} \ln h(G(x)) d\ln g(x)$$
$$= c\int_{S^1} dx \ln H' \circ G(x) \partial_x \ln G'(x)$$
$$= c\int_{S^1} dx \ln H' \circ G(x)\, \frac{G''(x)}{G'(x)}. \quad (16)$$

Hence

$$c\beta(\zeta,\eta) = c\int_{S^1} \eta'(x)\zeta''(x)dx$$
$$= \frac{c}{2}\int_{S^1} dx(\eta'''(x)\zeta(x) - \zeta'''(x)\eta(x)). \quad (17)$$

It follows that the Lie algebra of the group $\mathrm{Diff}(S^1)\hat{}$ is the standard Virasoro algebra $\mathrm{Vect}(S^1)\hat{}$.

I shall next go to 3 + 1 dimensions. Let a and f denote the Abelian gauge potential one-form and the gauge field two-form coupled to the gravity, respectively. The relevant term for the mixed anomaly is

$$\Omega_6^{-1}(f,R) = cf\,\mathrm{tr}\, R^2 \quad (f = da), \quad (18)$$

where c is a normalization constant so that

$$\int_{S^6} \Omega_6^{-1}(f,R) = c\int_{S^6} f\,\mathrm{tr}\, R^2 = \mathbb{Z}. \quad (19)$$

Locally Ω_6^{-1} is exact,

$$\Omega_6^{-1}(f,R) = d\Omega_5^0(f,\Gamma) \equiv d\Omega_5^0(f,0),$$

and Ω_5^0 is the Chern–Simons five-form,

$$\Omega_5^0(f,0) = cf\wedge \mathrm{tr}(\Gamma\, d\Gamma + \tfrac{2}{3}\Gamma^3). \quad (20)$$

Taking Δ of Ω_5^0 gives

$$(\Delta\Omega_5^0)(f,0,1) \equiv \Omega_5^0(f,1) - \Omega_5^0(f,0)$$
$$= cf\wedge d\,\mathrm{tr}(\Gamma\, dg\, g^{-1})$$
$$- c\tfrac{1}{3}f\wedge \mathrm{tr}(dg\, g^{-1})^3. \quad (21)$$

Thus we have the Einstein (or gravitational) anomaly in four dimensions:

$$\Omega_4^1(f,0,1) = cf\wedge \mathrm{tr}(\Gamma\, dg\, g^{-1}) + cf\wedge H^{(2)}(g). \quad (22)$$

Now let D^4 be the four-disk in six dimensions. Then any smooth one-to-one map G: $D^4 \to D^4$ can be restricted to the one G: $\partial D^4 \to \partial D^4$, where ∂D^4 is the boundary of D^4. We shall consider the case $S^3 = \partial D^4$. Let G and H be two diffeomorphisms of D^4 such that $G \circ H$: $D^4 \to D^4$. We can define

$$\alpha^2 = \int_{D^4} (\Delta\Omega_4^1)(f,0,1,2)$$
$$= c\int_{D^4} f\wedge \{\mathrm{tr}(h(G(x))^{-1}\, dh(G(x)) dg(x) g(x)^{-1})$$
$$+ H^{(2)}(g(x)) + H^{(2)}(h(x))$$
$$+ H^{(2)}(g(x)h(G(x)))\}. \quad (23)$$

Again modulo an integer Z the function α^2 is independent of the chosen disk D^4, and α^2 obeys the two-cocycle condition.

In the finite vicinity of the identity $g = 1 + u(x)$, we have

$$\alpha^2 = c\int_{S^3} f\wedge \mathrm{tr}(\ln h(G(x)) d\ln g(x)). \quad (24)$$

It follows that $\alpha^2 = 0$ if the Abelian gauge field $f = 0$ on the three-dimensional space S^3. Except for the factor f, the Schwinger–Jackiw–Johnson term [(23) and (24)] is the same as in the two-dimensional case studied earlier.

Suppose now that the three-dimensional space $S^3 = \partial D^4$ factorizes to $S^2 \times S^1$, which corresponds to periodic boundary conditions with respect to the radial variable r for the field components. Following Mickelsson,[16] we can interpret a point on S^2 as giving the polar angles in the physical three-space. If now f_{ij} is the field of a monopole, $f=\sin^2\theta\, d\theta\wedge d\phi$, then

$$\alpha^2(G,H)=c\int_{S^2\times D^2}\sin^2\theta\, d\theta\wedge d\phi \wedge\{\mathrm{tr}(h(G(x))^{-1}dh(G(x))dg(x)g(x)^{-1}) + H^{(2)}(g(x))+H^{(2)}(h(x)) + H^{(2)}(g(x)h(G(x)))\}, \quad (25)$$

where, as before, the boundary of D^2 is S^1. Equation (24) becomes

$$\alpha^2(G,H)=c\int_{S^2\times S^1}(\ln h(G(r,\hat{r}))\partial_r \ln g(r,\hat{r})) \times\sin^2\theta\, d\theta\wedge d\phi\wedge dr, \quad (26)$$

where $\hat{r}$ is the unit vector determined by the angles (θ,ϕ). It follows that

$$c\beta(\zeta,\eta)=c\int_{S^2\times S^1}(\partial_r\eta(r,\hat{r})\partial_r^2\zeta(r,\hat{r}))\sin^2\theta\, d\theta\wedge d\phi\wedge dr$$
$$=\frac{c}{2}\int_{S^2\times S^1}(\partial_r^3\eta(r,\hat{r})\zeta(r,\hat{r})-\partial_r^3\zeta(r,\hat{r})\eta(r,\hat{r})) \times\sin^2\theta\, d\theta\wedge d\phi\wedge dr, \quad (27)$$

which is just the arithmetic mean value over S^2 of the one-dimensional Virasoro algebra central extension

$$\frac{c}{2}\int_{S^1}(\eta'''(r)\zeta(r)-\zeta'''(r)\eta(r))dr. \quad (28)$$

Now we can define the Virasoro algebra and group on $S^2\times S^1$. Let $\mathrm{Diff}(S^2\times S^1)$ be the diffeomorphism group of smooth one-to-one maps $S^2\times S^1\to S^2\times S^1$ and $\mathrm{Lie}(\mathrm{Diff}(S^2\times S^1))\equiv\mathrm{Vect}(S^2\times S^1)$ its Lie algebra. We define the Virasoro algebra $\mathrm{Vect}(S^2\times S^1)^\wedge=\mathrm{Vect}(S^2\times S^1)\oplus i\,\mathbb{R}$ by the commutator

$$\left[\zeta(r,\hat{r})\frac{\partial}{\partial r},\eta(r,\hat{r})\frac{\partial}{\partial r}\right] =(\partial_r\zeta(r,\hat{r})\eta(r,\hat{r})-\partial_r\eta(r,\hat{r})\zeta(r,\hat{r}))\frac{\partial}{\partial r}+ic\beta(\zeta,\eta), \quad (29)$$

where $\zeta\,\partial/\partial r$, $\eta\,\partial/\partial r\in\mathrm{Vect}(S^2\times S^1)$ and $c\beta(\zeta,\eta)$ is given by (27). The corresponding Virasoro group $\mathrm{Diff}(S^2\times S^1)^\wedge=\mathrm{Diff}(S^2\times S^1)\times S^1$ is defined by the following multiplication law:

$$(G,\lambda)(H,\mu)=(G\circ H,\lambda\mu e^{i2\pi\alpha^2(G,H)}), \quad (30)$$

where $G,H\in\mathrm{Diff}(S^2\times S^1)$, $\lambda,\mu\in S^1$ (circle group), and α^2 is given by (25). It is not hard to see that the multiplication law (30) associates since the function α^2 obeys the two-cocycle condition. Equation (30) is seen to be a direct generalization of the one-dimensional Virasoro group.

ACKNOWLEDGMENTS

I would like to thank Professor B. Y. Hou and Professor P. Wang for a careful reading of the manuscript and helpful discussions.

This work was supported in part by the National Science Foundation of the People's Republic of China.

[1]W. B. Bardeen and B. Zumino, Nucl. Phys. B **244**, 421 (1984); L. Alvarez-Gaume and E. Witten, *ibid.* **234**, 269 (1983).
[2]O. Alvarez, I. M. Singer, and B. Zumino, Commun. Math. Phys. **96**, 409 (1984).
[3]L. Alvarez-Gaume and P. Ginsparg, Ann. Phys. (NY) **161**, 423 (1985).
[4]P. Goddard and D. Olive, Int. J. Mod. Phys. A **1**, 303 (1986).
[5]P. Goddard, DAMTP 85/7 (unpublished).
[6]G. Segal, Commun. Math. Phys. **80**, 301 (1981).
[7]R. Jackiw, TASI-86 Summer School, Santa Cruz, CA, CTP#1375, 1986.
[8]R. Floreanini and R. Jackiw, Phys. Lett. B **175**, 428 (1986).
[9]R. Floreanini, CTP#1370, 1986.
[10]R. Floreanini, C. T. Hill, and R. Jackiw, CTP#1412, 1986.
[11]Y. Z. Zhang, J. Phys. A: Math. Gen. **20**, 3099 (1987).
[12]L. D. Faddeev, Phys. Lett. B **145**, 81 (1984).
[13]C. Cronstrom and J. Mickelsson, J. Math. Phys. **24**, 2528 (1983).
[14]B. Y. Hou and P. Wang, Lett. Math. Phys. **11**, 179 (1986).
[15]Y.-Z. Zhang, Lett. Math. Phys. (in press); B. Y. Hou and Y. Z. Zhang, J. Phys. A: Math. Gen. (in press).
[16]J. Mickelsson, Phys. Rev. D **32**, 436 (1985).

$N = 4$ irreducible scalar superfields

S. V. Ketov and B. B. Lokhvitsky
Institute of High Current Electronics of the USSR Academy of Sciences Siberian Division, pr. Akademichesky, 4, 634055, Tomsk, USSR

(Received 14 August 1987; accepted for publication 21 October 1987)

The $N = 4$ extended supersymmetry in four-dimensional space-time is considered from the viewpoint of the two included $N = 2$ extended supersymmetries. Using the results for the structure of the irreducible superfields in $N = 2$ superspace obtained earlier the task of explicit construction of $N = 4$ irreducible superfields with $SU(2) \times SU(2)$ internal symmetry in $N = 2 \times 2$ superspace is solved. The decomposition of the $N = 4$ scalar superfield into $N = 4$ irreducible ones is described and the rules for collecting the irreducible scalar $N = 2 \times 2$ superfields with $SU(2) \times SU(2)$ internal symmetry into $N = 4$ irreducible scalar superfields with $SU(4)$ internal symmetry are set. To illustrate this method a simple consistent way of constructing the linearized off-shell $N = 4$ conformal supergravity in $N = 4$ superspace is shown.

I. INTRODUCTION

Over the last 10–15 years, a great deal of information has been obtained on supersymmetry and supergravity. The superspace formalism originally invented by Salam and Strathdee,[1] which proved to be adequate for the description of $N = 1$ supersymmetric theories, turns out to be very complicated in the extended supersymmetry case. As is well known, the reason for this is the increasing number of nonphysical components in a superfield as the number of anticommuting coordinates increases. The mass dimension of the superspace measure also increases while the dimensions of the physical fields remain unchanged. As a result, superspace formulations of extended theories are notoriously difficult to construct.

Recent developments in extended supersymmetric theories concerning the invention of the harmonic superspace approach by Galperin *et al.*[2] will possibly solve many of the problems in extended supersymmetry. Indeed, the success of this approach in describing $N = 2$ and $N = 3$ supersymmetric models is impressive. At the same time, the preference for $N = 4$ harmonic superspace over the usual $N = 4$ superspace is not obvious at all. There are some indications[3] that the harmonic superspace approach is unsufficient in its present form to solve the outstanding problems concerning the off-shell description of $N = 4$ supersymmetric field theories. Clearly, $N = 4$ supersymmetry is the one of special interest: it corresponds to $N = 1$ supersymmetry in the ten-dimensional space-time, which is relevant for superstrings: it is the maximally extended supersymmetry in the four-dimensional space-time, which can be incorporated in the construction of various off-shell $N = 4$ theories, e.g., $N = 4$ conformal supergravities.

This explains our interest in the investigation of the irreducible superfields in $N = 4$ extended superspace. The $N = 4$ supersymmetry was discussed in many works from various points of view.[4–15] In this work we develop our method, which was previously used for explicit construction of the irreducible $N = 2$ superfields and the analysis of the structure of $N = 2$ supersymmetric models.[16]

To give an example, we briefly summarize the decomposition of a general $N = 2$ scalar superfield into the irreducible ones by means of our method. According to the well-known results,[1] a general $N = 1$ scalar superfield decomposes into the sum of chiral, antichiral, and linear superfields. The $N = 1$ superprojectors are given by[1]

$$E_+ = -(1/4\Box)\bar{D}^2D^2, \quad E_- = -(1/4\Box)D^2\bar{D}^2, \quad E_1 = (1/2\Box)D^\alpha\bar{D}^2D_\alpha = (1/2\Box)\bar{D}_{\dot{\alpha}}D^2\bar{D}^{\dot{\alpha}}. \tag{1.1}$$

Consider now a general scalar $N = 2$ superfield, which depends on two sets of spinor anticommuting coordinates, each one being the set of the anticommuting coordinates for $N = 1$ superspace. Taking into account the decomposition (1.1), we can easily define $3 \times 3 = 9$ (maybe reducible) $SO(2)$-extended superfields in $N = 2$ superspace. One can show that all these superfields are, in fact, irreducible with the only exception being the last one $\phi_{1,1}$, which appears to be the sum of two irreducible superfields. At the second stage, arranging the anticommuting coordinates and the superspace covariant derivatives into fundamental $SU(2)$ doublets, it is not difficult to unite these $9 + 1 = 10$ $SO(2)$-irreducible superfields (or corresponding superprojectors) into the $SU(2)$ ones. Extensively using the information about $N = 1$ superfields,[1] we have explicitly constructed all irreducible $SO(2)$- and $SU(2)$-extended scalar superfields, $N = 2$ supersymmetry transformation laws for independent components, superprojectors, and invariant quadratic Lagrangians[16,17] and have generalized our consideration to $N = 2$ irreducible superfields with an external index.[17]

In this work we follow a similar approach, but apply it to $N = 4$ supersymmetry, which is considered from the viewpoint of the two included $N = 2$ supersymmetries. The preliminaries concerning the relevant facts about $N = 2$ and $N = 4$ superspaces are cited in Sec. II. Our conventions and notation can also be found in Sec. II. Section III is devoted to the explicit construction of the irreducible superprojectors in $N = 2 \times 2$ and $N = 4$ superspaces [in the discussion concerning the $N = 4$ superfields with $SU(2) \times SU(2)$ internal symmetry we prefer to use the term "$N = 2 \times 2$ superspace," while the usual term "$N = 4$ superspace" will be referred to the superfields with $SU(4)$ internal symmetry]. In Sec. IV

 0022-2488/88/051244-09$02.50

we apply our method to the construction of the particular $N=2\times 2$ superfield with $128+128$ independent components. As a by-product, the linearized off-shell $N=4$ conformal supergravity is formulated in Sec. V. Our conclusions are summarized in Sec. VI.

II. PRELIMINARIES CONCERNING THE $N=2$ AND $N=4$ SUPERSPACES; OUR NOTATION AND CONVENTIONS

Our conventions for the $d=3+1$ space-time signature and for the Levi-Civita symbol $\epsilon_{\mu\nu\lambda\rho}$ are

$$\eta_{\mu\nu}=\mathrm{diag}(+\,-\,-\,-),\quad \epsilon_{0123}=+1. \tag{2.1}$$

The conventions for 4×4 Dirac gamma matrices are

$$\begin{aligned}&\{\gamma_\mu,\gamma_\nu\}=2\eta_{\mu\nu},\\ &\gamma_5=\gamma_0\gamma_1\gamma_2\gamma_3,\quad \gamma_5^2=-1,\quad \hat{\partial}=\gamma_\mu\,\partial_\mu,\\ &C^{-1}\gamma_\mu C=-\gamma_\mu^T,\quad C^T=-C,\end{aligned} \tag{2.2}$$

where C is the charge conjugation matrix.

The convenient realization of Dirac gamma matrices, which is appropriate for the two-component formalism, is given by

$$\begin{aligned}&\gamma_0=\begin{pmatrix}0 & I_2\\ I_2 & 0\end{pmatrix},\quad \gamma_a=\begin{pmatrix}0 & \sigma_a\\ -\sigma_a & 0\end{pmatrix},\\ &i\gamma_5=\begin{pmatrix}I_2 & 0\\ 0 & -I_2\end{pmatrix},\end{aligned} \tag{2.3}$$

where σ_a $(a=1,2,3)$ are the usual Pauli matrices. The two-component formalism for spinors is based on the relations

$$\gamma_\mu=\begin{pmatrix}0 & \sigma_\mu\\ \tilde{\sigma}_\mu & 0\end{pmatrix},\quad \psi=\begin{pmatrix}\psi_\alpha\\ \bar{\psi}^{\dot\alpha}\end{pmatrix},\quad \bar{\psi}=(\psi^\alpha,\bar{\psi}_{\dot\alpha}) \tag{2.4}$$

(for details, see, e.g., Ref. 18), where ψ is any Majorana spinor in $d=4$: $\psi=C\bar{\psi}^T$. Explicitly,

$$\begin{aligned}&\sigma_{\mu\,\alpha\dot\beta}=(1,\sigma_a)_{\alpha\dot\beta},\quad \tilde{\sigma}_\mu^{\dot\alpha\beta}=(1,-\sigma_a)^{\dot\alpha\beta},\\ &\alpha=1,2,\quad \dot\alpha=\dot1,\dot2,\\ &\sigma_\mu\tilde{\sigma}_\nu+\sigma_\nu\tilde{\sigma}_\mu=2\eta_{\mu\nu},\quad \tilde{\sigma}_\mu\sigma_\nu+\tilde{\sigma}_\nu\sigma_\mu=2\eta_{\mu\nu},\\ &\partial_{\alpha\dot\alpha}=\sigma_{\mu\,\alpha\dot\alpha}\,\partial_\mu,\quad \partial^{\alpha\dot\alpha}=\tilde{\sigma}_\mu^{\dot\alpha\alpha}\,\partial_\mu,\\ &\partial_{\alpha\dot\alpha}\partial^{\dot\alpha\beta}=\delta_\alpha^\beta\Box,\quad \partial^{\dot\alpha\alpha}\partial_{\alpha\dot\beta}=\delta_{\dot\beta}^{\dot\alpha}\Box,\\ &\tilde{\sigma}_\mu^{\dot\alpha\alpha}=\epsilon^{\alpha\beta}\epsilon^{\dot\alpha\dot\beta}\sigma_{\mu\,\beta\dot\beta},\quad \sigma_{\mu\,\alpha\dot\alpha}=\epsilon_{\alpha\beta}\epsilon_{\dot\alpha\dot\beta}\tilde{\sigma}_\mu^{\dot\beta\beta},\\ &\mathrm{tr}(\sigma_\mu\tilde{\sigma}_\nu)=2\eta_{\mu\nu}.\end{aligned} \tag{2.5}$$

Our conventions for the two-component spinors are

$$\begin{aligned}&\psi_\alpha=\epsilon_{\alpha\beta}\psi^\beta,\quad \psi^\beta=\epsilon^{\beta\gamma}\psi_\gamma,\\ &\bar{\psi}_{\dot\alpha}=\epsilon_{\dot\alpha\dot\beta}\bar{\psi}^{\dot\beta},\quad \bar{\psi}^{\dot\beta}=\epsilon^{\dot\beta\dot\alpha}\bar{\psi}_{\dot\alpha},\\ &\epsilon_{\alpha\beta}\epsilon^{\beta\gamma}=\delta_\alpha^\gamma,\quad \epsilon_{\dot\alpha\dot\beta}\epsilon^{\dot\beta\dot\gamma}=\delta_{\dot\alpha}^{\dot\gamma},\\ &\epsilon_{12}=\epsilon^{21}=\epsilon^{\dot2\dot1}=\epsilon_{\dot1\dot2}=-1.\end{aligned} \tag{2.6}$$

To simplify the notation, Lorentz SL$(2,C)$ indices are sometimes suppressed in their "natural" position, for example,

$$\begin{aligned}&(\theta\sigma_\mu\bar\eta)\equiv\theta^\alpha\sigma_{\mu\,\alpha\dot\beta}\bar\eta^{\dot\beta}=-(\bar\eta\tilde\sigma_\mu\theta),\\ &(\theta\sigma_{\mu\nu}\eta)\equiv\theta^\alpha(\sigma_{\mu\nu})_\alpha^{\ \beta}\eta_\beta=-(\eta\sigma_{\mu\nu}\theta).\end{aligned} \tag{2.7}$$

To designate Lorentz indices, Greek letters are used throughout this paper. These indices are raised and lowered with ϵ symbols as in (2.6). For SU(2) indices we use lower case Latin letters, whereas capital Latin letters are used for SU(4) indices. Commas separate the indices referred to SU(2)$\times$SU(2), viz.,

$$X^A=(X^{i\cdot},X^{\cdot i}),\quad A=1,\ldots,4,\quad i=1,2. \tag{2.8}$$

The totally antisymmetric SU(4)-invariant ϵ symbol is normalized as follows:

$$\epsilon^{1234}=-\epsilon_{1234}=1. \tag{2.9}$$

With this normalization conventions the following relations are valid:

$$\begin{aligned}&\epsilon^{ij}\epsilon_{ij}=-2,\quad \epsilon^{ij}\epsilon_{jk}=\delta_k^i,\quad \epsilon^{ABCD}\epsilon_{ABCD}=-24,\\ &\epsilon^{ABCD}\epsilon_{ABCE}=-6\delta_E^D,\quad \epsilon^{ABCD}\epsilon_{ABMN}=-2\delta_{[M}^C\,\delta_{N\,]}^D,\\ &X^i=\epsilon^{ij}X_j,\quad X_i=\epsilon_{ij}X^j,\quad X^A=(1/\sqrt6)\epsilon^{ABCD}X_{BCD},\\ &X_{BCD}=-(1/\sqrt6)\epsilon_{ABCD}X^A.\end{aligned} \tag{2.10}$$

The symmetrization and antisymmetrization are defined without normalizing factors, viz.,

$$\begin{aligned}&X_{[AB\,]}=X_{AB}-X_{BA},\quad X_{(AB)}=X_{\{AB\}}=X_{AB}+X_{BA},\\ &X_{(ABC)}=X_{ABC}+X_{CAB}+X_{BCA},\\ &X_{\{A_1\cdots A_n\}}=\sum_{\substack{\text{all permutations}\\(i_1\cdots i_n)}}X_{A_{i_1}\cdots A_{i_n}}.\end{aligned} \tag{2.11}$$

The products of the Grassman anticommuting coordinates of $N=2$ superspace are given by

$$\begin{aligned}&\theta_{ij}=\theta_i^\alpha\theta_{\alpha j},\quad \theta_{\alpha\beta}=\theta_{\alpha i}\theta_\beta^i,\quad \theta_{ij}\theta_{\alpha\beta}=0,\\ &(\theta^3)_\alpha^i=\tfrac13\theta^{ij}\theta_{j\alpha}=\tfrac13\theta_{\alpha\beta}\theta^{\beta i},\\ &\theta^4=\tfrac1{12}\theta_{ij}\theta^{ij}=-\tfrac1{12}\theta_{\alpha\beta}\theta^{\alpha\beta}=\tfrac14\theta_i^\alpha(\theta^3)_\alpha^i.\end{aligned} \tag{2.12}$$

Similar relations can be found for $N=2$ superspace covariant derivatives. The analogous formulas for the $\bar\theta$'s ($\bar D$'s) easily follow by complex conjugation.

The products of the anticommuting coordinates of $N=4$ superspace are defined in a similar fashion:

$$\begin{aligned}&\theta_\alpha^A=(\theta_\alpha^{i\cdot},\theta_\alpha^{\cdot i}),\quad \theta^{AB}=\tfrac12\theta^{A\alpha}\theta_\alpha^B,\\ &\theta_{\alpha\beta}^{[AB]}\equiv\tfrac12\theta_\alpha^{[A}\theta_\beta^{B]}=\tfrac12\theta_{(\alpha}^A\theta_{\beta)}^B,\\ &(\theta^3)_{\alpha\beta\gamma D}=-\tfrac16\epsilon_{ABCD}\theta_\alpha^A\theta_\beta^B\theta_\gamma^C,\\ &(\theta^3)_\alpha^{[AB]C}=-\tfrac13\theta_{\alpha\beta}^{[AB]}\theta^{\beta C}=\tfrac13\theta_\alpha^{[A}\theta^{B]C},\\ &(\theta^4)_{\alpha\beta\gamma\delta}=-\tfrac1{24}\epsilon_{ABCD}\theta_\alpha^A\theta_\beta^B\theta_\gamma^C\theta_\delta^D,\\ &(\theta^4)_{B\alpha\beta}^A=(\theta^3)_{\alpha\beta\gamma B}\theta^{\gamma A}=-\tfrac13\epsilon_{BCDE}\theta^{AC}\theta_{\alpha\beta}^{[DE]},\\ &(\theta^4)_{[CD]}^{[AB]}=-\tfrac1{12}\epsilon_{CDEF}\theta^{[AB]\alpha\beta}\theta_{\alpha\beta}^{[EF]}\\ &\qquad=\tfrac14\epsilon_{CDEF}(\theta^3)_\alpha^{[EF]A}\theta^{B\alpha}\\ &\qquad=\tfrac14\epsilon_{CDEF}(\theta^3)_\alpha^{[AB]E}\theta^{F\alpha}\\ &\qquad=\tfrac16\epsilon_{CDEF}\theta^{[AE}\theta^{B]F}.\end{aligned} \tag{2.13}$$

In actual calculations the reduction formulas for the products of anticommuting coordinates are needed. In $N=2$ superspace they are given by

$$
\begin{aligned}
&\theta^i_\alpha\theta^j_\beta=\tfrac{1}{2}\epsilon_{\alpha\beta}\theta^{ij}+\tfrac{1}{2}\epsilon^{ij}\theta_{\alpha\beta},\\
&\theta^i_\alpha\theta^j_\beta\theta^k_\gamma=\tfrac{1}{2}\epsilon_{\alpha\beta}\epsilon^{k(i}(\theta^3)^{j)}_\gamma-\tfrac{1}{2}\epsilon^{ij}\epsilon_{\gamma(\alpha}(\theta^3)^k_{\beta)},\\
&\theta^i_\alpha\theta^j_\beta\theta^k_\gamma\theta^l_\delta=\tfrac{1}{2}(\epsilon^{ij}\epsilon^{kl}\epsilon_{\alpha(\gamma}\epsilon_{\delta)\beta}-\epsilon_{\alpha\beta}\epsilon_{\gamma\delta}\epsilon^{i(k}\epsilon^{l)j})\theta^4.
\end{aligned}
\tag{2.14}
$$

The reduction formulas in $N=4$ superspace are of the form

$$
\begin{aligned}
\theta^A_\alpha\theta^B_\beta&=\epsilon_{\alpha\beta}\theta^{AB}+\theta^{[AB]}_{\alpha\beta},\\
\theta^A_\alpha\theta^B_\beta\theta^C_\gamma&=\epsilon^{ABCD}(\theta^3)_{\alpha\beta\gamma D}+\tfrac{1}{3}((\theta^3)^{[AB]C}_{(\alpha}\epsilon_{\beta)\gamma}\\
&\quad+(\theta^3)^{[BC]A}_{(\beta}\epsilon_{\gamma)\alpha}+(\theta^3)^{[CA]B}_{(\gamma}\epsilon_{\alpha)\beta}),\\
\theta^A_\alpha\theta^B_\beta\theta^C_\gamma\theta^D_\delta&=\epsilon^{ABCD}(\theta^4)_{\alpha\beta\gamma\delta}\\
&\quad+\tfrac{1}{4}\{\epsilon^{EAB(C}\epsilon_{\gamma\delta}(\theta^4)^{D)}_{E\alpha\beta}-\epsilon^{EDA(B}\epsilon_{\beta\gamma}(\theta^4)^{C)}_{E\delta\alpha}\\
&\quad+\epsilon^{ECD(A}\epsilon_{\alpha\beta}(\theta^4)^{B)}_{E\gamma\delta}-\epsilon^{EBC(D}\epsilon_{\delta\alpha}(\theta^4)^{A)}_{E\beta\gamma}\}\\
&\quad+\tfrac{1}{4}\{\epsilon^{EFBC}\epsilon_{\alpha(\beta}\epsilon_{\gamma)\delta}(\theta^4)^{[DA]}_{[EF]}\\
&\quad-\epsilon^{EFCD}\epsilon_{\beta(\gamma}\epsilon_{\delta)\alpha}(\theta^4)^{[AB]}_{[EF]}\\
&\quad+\epsilon^{EFDA}\epsilon_{\gamma(\delta}\epsilon_{\alpha)\beta}(\theta^4)^{[BC]}_{[EF]}\\
&\quad-\epsilon^{EFAB}\epsilon_{\delta(\alpha}\epsilon_{\beta)\gamma}(\theta^4)^{[CD]}_{[EF]}\}.
\end{aligned}
\tag{2.15}
$$

Thus we conclude that the SU(2)-irreps θ^α_i, $\theta_{\alpha\beta}$, θ_{ij}, and $(\theta^3)^i_\alpha$ realize the following representations of SU(2): 2, 1, 3, and 2, respectively, whereas the SU(4)-irreps θ^A_α, θ^{AB}, $\theta^{[AB]}_{\alpha\beta}$, $(\theta^3)^{\alpha\beta\gamma}_D$, $(\theta^3)^{[AB]C}_\alpha$, $(\theta^4)_{\alpha\beta\gamma\delta}$, $(\theta^4)^A_{B\alpha\beta}$, and $(\theta^4)^{[AB]}_{[CD]}$ transform as 4, 10, 6, $\bar{4}$, 20, 1, 15, and 20′ of SU(4), respectively. Consequently,

$$
\epsilon_{ABCD}(\theta^3)^{[AB]C}_\alpha=0,\quad(\theta^4)^A_{A\,\alpha\beta}=(\theta^4)^{[AB]}_{[AC]}=0.
\tag{2.16}
$$

Our normalizations of the superintegration measure in $N=2$ and $N=4$ superspaces are given by

$$
\begin{aligned}
&\int d^4\theta\,\theta^4=\int d^4\bar\theta\,\bar\theta^4=1,\\
&\int d^8\theta\,\theta^8=\int d^8\bar\theta\,\bar\theta^8=1.
\end{aligned}
\tag{2.17}
$$

Finally, the covariant derivatives in $N=2$ superspace $(D^i_\alpha,\bar D^{\dot\alpha}_i)$ and in $N=4$ superspace $(D^A_\alpha,\bar D^{\dot\alpha}_A)$ satisfy the algebras

$$
\{D^i_\alpha,D^j_\beta\}=\{\bar D^{\dot\alpha}_i,\bar D^{\dot\beta}_j\}=0,\quad\{D^i_\alpha,\bar D_{\dot\alpha j}\}=i\,\partial_{\alpha\dot\alpha}\delta^i_j,
\tag{2.18}
$$

and

$$
\{D^A_\alpha,D^B_\beta\}=\{\bar D^{\dot\alpha}_A,\bar D^{\dot\beta}_B\}=0,\quad\{D^A_\alpha,\bar D_{\dot\alpha B}\}=i\,\partial_{\alpha\dot\alpha}\delta^A_B,
\tag{2.19}
$$

respectively. All Grassman derivatives are taken to be left ones.

A general $N=2$ scalar superfield decomposes into the sum of six $N=2$ irreducible ones with U(2) internal symmetry.[8,12,16] The $N=2$ superprojectors are given by[8,16]

$$
\begin{aligned}
\Pi_0&\equiv\Pi_{\langle0,0,0\rangle}=(1/\Box^2)\bar D^4D^4,\\
\Pi_1&\equiv\Pi_{\langle1/2,1/2,1\rangle}=(1/\Box^2)D^{\alpha i}\bar D^4(D^3)_{\alpha i},\\
\Pi_2&\equiv\Pi_{\langle0,1,2\rangle}=(1/4\Box^2)D_{ij}\bar D^4D^{ij},\\
\Pi_3&\equiv\Pi_{\langle1,0,2\rangle}=-(1/4\Box^2)D^{\alpha\beta}\bar D^4D_{\alpha\beta},\\
\Pi_4&\equiv\Pi_{\langle1/2,1/2,3\rangle}=(1/\Box^2)(D^3)^\alpha_i\bar D^4D^i_\alpha,\\
\Pi_5&\equiv\Pi_{\langle0,0,4\rangle}=(1/\Box^2)D^4\bar D^4,
\end{aligned}
\tag{2.20}
$$

where the numbers in brackets indicate the quantum numbers of the three Casimir operators of $N=2$ superalgebra: superspin, superisospin, and $N=2$ supercharge, respectively.[16] We display here only one real superfield ($N=2$ conformal supercurrent) $\phi_{\langle1,0,2\rangle}$, which will be used in its explicit form in our further calculations.

The information about the component structure of $\phi_{\langle1,0,2\rangle}$ is collected in Table I.

The superfield itself reads as follows:

$$
\begin{aligned}
\phi_{\langle1,0,2\rangle}&=D+(\theta_i\sigma_\mu\bar\theta_j)I^{ij}_\mu+(\theta_i\sigma_\mu\bar\theta^i)I^{(5)}_\mu\\
&\quad+\tfrac{1}{24}(\theta_i\sigma_\mu\bar\theta^i)(\theta_j\sigma_\rho\bar\theta^j)(\Box\eta_{\mu\rho}-\partial_\mu\,\partial_\rho)D-\tfrac{1}{16}\theta^{\alpha\beta}\sigma_{\mu\,\alpha\dot\alpha}\sigma_{\nu\,\beta\dot\beta}\bar\theta^{\dot\alpha\dot\beta}T^{(2)}_{\mu\nu}+\tfrac{1}{4}(\theta^{3i}\sigma_\mu\bar\theta^3_i)\Box I^{(5)}_\mu\\
&\quad+\tfrac{1}{4}(\theta^3_i\sigma_\mu\bar\theta^3_j)\Box I^{ij}_\mu+\tfrac{1}{16}\theta^4\bar\theta^4\Box^2D+\{\theta_i\lambda^i+\tfrac{1}{2}\theta^{\alpha\beta}G_{\alpha\beta}+\tfrac{1}{4}\theta^{\alpha\beta}(\sigma_\mu\bar\theta_i)_\alpha Z^{(3/2)i}_{\mu\beta}+\tfrac{1}{4}\theta_{ij}(\bar\theta^i i\,\partial\lambda^j)\\
&\quad-(i/24)\theta^{\alpha\beta}(\sigma_\mu\bar\theta_i)_\alpha(\sigma_{\mu\rho}\,\partial_\rho\lambda^i)_\beta+\tfrac{1}{2}(\theta^3)^\alpha_i(\bar\theta^i i\partial)^\beta G_{\beta\alpha}+(i/8)\theta^\alpha_\beta\bar\theta_{ij}(\sigma_{\mu\rho}\,\partial_\rho)^\beta_\alpha I^{ij}_\mu+\tfrac{1}{8}\bar\theta_{ij}(\theta^{3i}\Box\lambda^j)\\
&\quad+\tfrac{1}{8}\bar\theta_{\dot\alpha\dot\beta}\tilde\sigma^{\dot\beta\alpha}_\mu(\theta^3)^j_\alpha i\,\partial^{\dot\alpha\beta}(Z^{(3/2)}_{\mu\beta j}-(i/6)[\sigma_{\mu\rho}\,\partial_\rho\lambda_j]_\beta)-\tfrac{1}{8}\theta^4\bar\theta_{\dot\alpha\dot\beta}\,\partial^{\alpha\dot\alpha}\,\partial^{\beta\dot\beta}G_{\alpha\beta}+\tfrac{1}{8}\theta^4(\bar\theta^3_j i\,\partial\lambda^j)+\text{H.c.}\}.
\end{aligned}
\tag{2.21}
$$

Consequently, the $N=2$ supersymmetry transformation laws are given by

$$
\begin{aligned}
\delta D&=\epsilon_i\lambda^i+\text{H.c.},\\
\delta\lambda^i_\alpha&=-(i/2)(\partial\bar\epsilon^i)_\alpha D+\epsilon^{\beta i}G_{\alpha\beta}\\
&\quad+(\sigma_\mu\bar\epsilon^i)_\alpha I^{(5)}_\mu+(\sigma_\mu\bar\epsilon_j)_\alpha I^{ij}_\mu,\\
\delta G_{\alpha\beta}&=\tfrac{1}{4}(\sigma_\mu\bar\epsilon_i)_{(\alpha}Z^{(3/2)i}_{\beta)\mu}+\tfrac{1}{3}(\partial\bar\epsilon_i)_{(\alpha}\lambda^i_{\beta)},\\
\delta I^{(5)}_\mu&=-\tfrac{1}{4}(\epsilon_iZ^{(3/2)i}_\mu)+(i/6)(\epsilon_i\sigma_{\mu\rho}\,\partial_\rho\lambda^i)+\text{H.c.},\\
\delta I^{ij}_\mu&=\tfrac{1}{8}(\sigma_\mu\tilde\sigma_\nu\epsilon^{(i})_\alpha Z^{(3/2)j)\alpha}_\nu-(i/6)(\sigma_{\mu\nu}\,\partial_\nu\epsilon^{(i})_\alpha\lambda^{j)\alpha}\\
&\quad+\text{H.c.},
\end{aligned}
\tag{2.22}
$$

$$
\begin{aligned}
\delta Z^{(3/2)i}_{\mu\alpha}&=-(i/2)(\epsilon^i\sigma_{\mu\nu}\,\partial_\nu)^\beta G_{\beta\alpha}\\
&\quad+(i/6)\epsilon^{\beta i}(\sigma_{\mu\nu}\,\partial_\nu)^\gamma_\alpha G_{\gamma\beta}\\
&\quad-\tfrac{1}{2}(\sigma_\rho\bar\epsilon^i)_\alpha T^{(2)}_{\mu\rho}+(2i/3)\partial_{\alpha\dot\beta}\\
&\quad\times(I^{(5)}_\mu\bar\epsilon^{\dot\beta i}-2I^{ij}_\mu\bar\epsilon^{\dot\beta}_j)\\
&\quad-\tfrac{1}{3}\epsilon_{\mu\nu\lambda\rho}\,\partial_\nu(\sigma_\rho)_{\alpha\dot\beta}(I^{(5)}_\lambda\bar\epsilon^{\dot\beta i}-2I^{ij}_\lambda\bar\epsilon^{\dot\beta}_j),\\
\delta T^{(2)}_{\mu\rho}&=-(i/4)(\epsilon_i\sigma_{(\mu\nu}\,\partial_\nu Z^{(3/2)i}_{\rho)})+\text{H.c.}
\end{aligned}
$$

Having obtained the superfield, we can therefore calculate the invariant quadratic Lagrangian in a straightforward way as follows:

TABLE I. The component structure of $\phi_{(1,0,2)}$. The highest spin is 2. The highest isospin is 1. All fields are irreducible w.r.t. $SL(2,C)\times SU(2)$. Altogether, there are 24 + 24 field components.

Components	Spin	SL(2,C) irrep	SU(2) irrep	Number of degrees of freedom	Young tableau
$T^{(2)}_{\mu\nu}$	2	(1,1)	1	5	$\begin{matrix}\boxed{\bullet}\boxed{\times}\\ \boxed{\times}\boxed{\bullet}\end{matrix}$
$Z^{(3/2)i}_{\mu\alpha}, \bar{Z}^{(3/2)\dot{\alpha}}_{\mu i}$	$\frac{3}{2}$	$(1,\frac{1}{2}) + (\frac{1}{2},1)$	2	8 + 8	$\begin{matrix}\boxed{\times}\boxed{\bullet}\\ \boxed{\bullet}\end{matrix}, \begin{matrix}\boxed{\bullet}\boxed{\times}\\ \boxed{\times}\end{matrix}$
$I^{(5)}_{\mu}$	1	$(\frac{1}{2},\frac{1}{2})$	1	3	$\begin{matrix}\boxed{\bullet}\\ \boxed{\times}\end{matrix}$
I^{ij}_{μ}	1	$(\frac{1}{2},\frac{1}{2})$	3	9	$\boxed{\bullet}\boxed{\times}$
$G_{\alpha\beta}, \bar{G}_{\dot{\alpha}\dot{\beta}}$	1	(1,0) + (0,1)	1	3 + 3	$\begin{matrix}\boxed{\bullet}\\ \boxed{\bullet}\end{matrix}, \begin{matrix}\boxed{\times}\\ \boxed{\times}\end{matrix}$
$\lambda^{i}_{\alpha}, \bar{\lambda}^{\dot{\alpha}}_{i}$	$\frac{1}{2}$	$(\frac{1}{2},0) + (0,\frac{1}{2})$	2	4 + 4	$\boxed{\bullet}, \boxed{\times}$
D	0	(0,0)	1	1	$\bullet$

$$\begin{aligned} L &= \int d^4\theta\, d^4\bar{\theta}\, \phi^2_{(1,0,2)} \\ &= \tfrac{1}{3} D\,\Box^2 D + \tfrac{4}{3}(\lambda_j i\,\partial\,\Box\bar{\lambda}^j) + 2G^{\alpha\beta}\,\partial_{\alpha\dot{\alpha}}\,\partial_{\beta\dot{\beta}}\bar{G}^{\dot{\alpha}\dot{\beta}} \\ &\quad + 2I^{(5)}_{\mu}\Box I^{(5)}_{\mu} - 2I^{ij}_{\mu}\Box I_{\mu ij} \\ &\quad - (Z^{(3/2)k}_{\mu} i\,\partial\bar{Z}^{(3/2)}_{\mu k}) + \tfrac{1}{4} T^{(2)}_{\mu\nu} T^{(2)}_{\mu\nu}. \end{aligned} \tag{2.23}$$

III. $N=2\times 2$ AND $N=4$ IRREDUCIBLE SUPERPROJECTORS

Let D^i_{α}, $\bar{D}^{\dot{\alpha}}_j$ be the covariant $N=2$ superspace derivatives w.r.t. the first $N=2$ supersymmetry, and let B^i_{α}, $\bar{B}^{\dot{\alpha}}_j$ be the ones w.r.t. the second $N=2$ supersymmetry. Both sets are contained in $N=4$ supersymmetry.

Taking direct products of six $N=2$ superprojectors (2.20) in terms of D-covariant derivatives to six $N=2$ superprojectors (2.20) in terms of B-covariant derivatives, we easily find $6\times 6=36$ (maybe reducible) superprojectors in $N=2\times 2$ superspace. The subsequent reduction of these $N=2\times 2$ superprojectors into $N=2\times 2$ irreducible ones is achieved solely by algebraic operations of symmetrization and antisymmetrization w.r.t. the Lorentz $SL(2,C)$ indices only. Then the symmetry properties of the internal symmetry indices are induced by the symmetry structure of Lorentz indices. As a result, we find 50 irreducible $N=2\times 2$ superprojectors which realize 50 irreducible $N=2\times 2$ scalar superfields with internal $SU(2)\times SU(2)$ symmetry by construction. Explicitly,

$$\Pi_{00} = (1/\Box^4)\bar{D}^4\bar{B}^4 D^4 B^4,$$
$$\Pi_{01} = -(1/\Box^4) B^{\alpha}_{i}\bar{D}^4\bar{B}^4 D^4 (B^3)^{i}_{\alpha},$$
$$\Pi_{02} = (1/4\Box^4) B_{ij}\bar{D}^4\bar{B}^4 D^4 B^{ij},$$
$$\Pi_{03} = -(1/4\Box^4) B^{\alpha\beta}\bar{D}^4\bar{B}^4 D^4 B_{\alpha\beta},$$
$$\Pi_{04} = (1/\Box^4)(B^3)^{\alpha}_{i}\bar{D}^4\bar{B}^4 D^4 B^{i}_{\alpha},$$
$$\Pi_{05} = (1/\Box^4) B^4\bar{D}^4\bar{B}^4 D^4,$$
$$\Pi_{10} = -(1/\Box^4) D^{\alpha}_{i}\bar{D}^4\bar{B}^4 (D^3)^{i}_{\alpha} B^4,$$
$$\Pi'_{11} = -(1/4\Box^4) D^{(\alpha}_{i} B^{\beta)}_{j}\bar{D}^4\bar{B}^4 (D^3)^{i}_{(\alpha}(B^3)^{j}_{\beta)},$$
$$\Pi''_{11} = -(1/2\Box^4) D^{\alpha}_{i} B_{j\alpha}\bar{D}^4\bar{B}^4 (D^3)^{i\beta}(B^3)^{j}_{\beta},$$
$$\Pi'_{12} = (1/16\Box^4) D^{(\alpha}_{i} B^{\beta)}_{j} B_{\beta k}\bar{D}^4\bar{B}^4 (D^3)^{i}_{(\alpha} B^{j}_{\gamma)} B^{\gamma k},$$
$$\Pi''_{12} = (1/16\Box^4) D^{\beta}_{i} B_{\beta j} B^{\alpha}_{k}\bar{D}^4\bar{B}^4 (D^3)^{i\gamma} B^{j}_{\gamma} B^{k}_{\alpha},$$
$$\Pi'_{13} = (1/36\Box^4) D^{(\alpha}_{i} B^{\beta\gamma)}\bar{D}^4\bar{B}^4 (D^3)^{i}_{(\alpha} B_{\beta\gamma)},$$
$$\Pi''_{13} = -(1/6\Box^4) D_{\alpha i} B^{\alpha\beta}\bar{D}^4\bar{B}^4 (D^3)^{i\gamma} B_{\beta\gamma},$$
$$\Pi'_{14} = (1/4\Box^4) D^{(\alpha}_{i}(B^3)^{\beta)}_{j}\bar{D}^4\bar{B}^4 (D^3)^{i}_{(\alpha} B^{j}_{\beta)},$$
$$\Pi''_{14} = (1/2\Box^4) D^{\alpha}_{i}(B^3)_{j\alpha}\bar{D}^4\bar{B}^4 (D^3)^{i\beta} B^{j}_{\beta},$$
$$\Pi_{15} = -(1/\Box^4) D^{\alpha}_{i} B^4\bar{D}^4\bar{B}^4 (D^3)^{i}_{\alpha},$$
$$\Pi_{20} = (1/4\Box^4) D_{ij}\bar{D}^4\bar{B}^4 D^{ij} B^4,$$
$$\Pi'_{21} = (1/16\Box^4) D_{i\beta} D^{(\beta}_{j} B^{\alpha)}_{k}\bar{D}^4\bar{B}^4 D^{\gamma i} D^{j}_{(\gamma}(B^3)^{k}_{\alpha)},$$
$$\Pi''_{21} = (1/16\Box^4) D^{\alpha}_{i} D^{\beta}_{j} B_{\beta k}\bar{D}^4\bar{B}^4 D^{i}_{\alpha} D^{\gamma j}(B^3)^{k}_{\gamma},$$
$$\Pi_{22} = (1/16\Box^4) D_{ij} B_{mn}\bar{D}^4\bar{B}^4 D^{ij} B^{mn},$$
$$\Pi_{23} = -(1/16\Box^4) D_{ij} B_{\alpha\beta}\bar{D}^4\bar{B}^4 D^{ij} B^{\alpha\beta},$$
$$\Pi'_{24} = -(1/16\Box^4) D_{i\beta} D^{(\beta}_{j}(B^3)^{\alpha)}_{k}\bar{D}^4\bar{B}^4 D^{\gamma i} D^{j}_{(\gamma} B^{k}_{\alpha)},$$

$$
\begin{aligned}
\Pi''_{24} &= -(1/16\Box^4)D^{\alpha}_{i}D^{\beta}_{j}(B^3)_{\beta k}\bar{D}^4\bar{B}^4D^{i}_{\alpha}D^{\gamma j}B^{k}_{\gamma},\\
\Pi_{25} &= (1/4\Box^4)D_{ij}B^4\bar{D}^4\bar{B}^4D^{ij},\\
\Pi_{30} &= -(1/4\Box^4)D^{\alpha\beta}\bar{D}^4\bar{B}^4D_{\alpha\beta}B^4, \qquad (3.1)\\
\Pi'_{31} &= (1/36\Box^4)D^{(\alpha\beta}B^{\gamma)}_{i}\bar{D}^4\bar{B}^4D_{(\alpha\beta}(B^3)^{i}_{\gamma)},\\
\Pi''_{31} &= -(1/6\Box^4)D^{\alpha\beta}B_{\beta i}\bar{D}^4\bar{B}^4D_{\alpha\gamma}(B^3)^{\gamma i},\\
\Pi_{32} &= -(1/16\Box^4)D^{\alpha\beta}B^{ij}\bar{D}^4\bar{B}^4D_{\alpha\beta}B_{ij},\\
\Pi'_{33} &= (1/16\Box^4)1/(4!)^2D^{\{\alpha\beta}B^{\gamma\delta\}}\bar{D}^4\bar{B}^4D_{\{\alpha\beta}B_{\gamma\delta\}},\\
\Pi''_{33} &= -(1/96\Box^4)D^{\gamma}_{(\alpha}B_{\beta)\gamma}\bar{D}^4\bar{B}^4D^{(\alpha}_{\gamma}B^{\beta)\delta},\\
\Pi'''_{33} &= (1/48\Box^4)D^{\alpha\beta}B_{\alpha\beta}\bar{D}^4\bar{B}^4D^{\gamma\delta}B_{\gamma\delta},\\
\Pi'_{34} &= -(1/36\Box^4)D^{(\alpha\beta}(B^3)^{\gamma)}_{i}\bar{D}^4\bar{B}^4D_{(\alpha\beta}B^{i}_{\gamma)},\\
\Pi''_{34} &= (1/6\Box^4)D^{\alpha\beta}(B^3)_{i\beta}\bar{D}^4\bar{B}^4D_{\alpha\gamma}B^{\gamma i},\\
\Pi_{35} &= -(1/4\Box^4)D^{\alpha\beta}B^4\bar{D}^4\bar{B}^4D_{\alpha\beta},\\
\Pi_{40} &= (1/\Box^4)(D^3)^{\alpha}_{i}\bar{D}^4\bar{B}^4D^{i}_{\alpha}B^4,\\
\Pi'_{41} &= (1/4\Box^4)(D^3)^{(\alpha}_{i}B^{\beta)}_{j}\bar{D}^4\bar{B}^4D^{i}_{(\alpha}(B^3)^{j}_{\beta)},\\
\Pi''_{41} &= (1/2\Box^4)(D^3)^{\alpha}_{i}B_{j\alpha}\bar{D}^4\bar{B}^4D^{\beta i}(B^3)^{j}_{\beta},\\
\Pi'_{42} &= -(1/16\Box^4)(D^3)^{(\alpha}_{i}B^{\beta)}_{j}B_{\beta k}\bar{D}^4\bar{B}^4D^{i}_{(\alpha}B^{j}_{\gamma)}B^{\gamma k},\\
\Pi''_{42} &= -(1/16\Box^4)(D^3)^{\beta}_{i}B_{\beta j}B^{\alpha}_{k}\bar{D}^4\bar{B}^4D^{\gamma i}B^{j}_{\gamma}B^{k}_{\alpha},\\
\Pi'_{43} &= -(1/36\Box^4)(D^3)^{(\alpha}_{i}B^{\beta\gamma)}\bar{D}^4\bar{B}^4D^{i}_{(\alpha}B_{\beta\gamma)},\\
\Pi''_{43} &= (1/6\Box^4)(D^3)_{i\alpha}B^{\alpha\beta}\bar{D}^4\bar{B}^4D^{i\gamma}B_{\beta\gamma},\\
\Pi'_{44} &= -(1/4\Box^4)(D^3)^{(\alpha}_{i}(B^3)^{\beta)}_{j}\bar{D}^4\bar{B}^4D^{i}_{(\alpha}B^{j}_{\beta)},\\
\Pi''_{44} &= -(1/2\Box^4)(D^3)^{\alpha}_{i}(B^3)_{j\alpha}\bar{D}^4\bar{B}^4D^{i\beta}B^{j}_{\beta},\\
\Pi_{45} &= (1/\Box^4)(D^3)^{\alpha}_{i}B^4\bar{D}^4\bar{B}^4D^{i}_{\alpha},\\
\Pi_{50} &= (1/\Box^4)D^4\bar{D}^4\bar{B}^4B^4,\\
\Pi_{51} &= -(1/\Box^4)D^4B^{\alpha}_{i}\bar{D}^4\bar{B}^4(B^3)^{i}_{\alpha},\\
\Pi_{52} &= (1/4\Box^4)D^4B_{ij}\bar{D}^4\bar{B}^4B^{ij},\\
\Pi_{53} &= -(1/4\Box^4)D^4B^{\alpha\beta}\bar{D}^4\bar{B}^4B_{\alpha\beta},\\
\Pi_{54} &= (1/\Box^4)D^4(B^3)^{\alpha}_{i}\bar{D}^4\bar{B}^4B^{i}_{\alpha},\\
\Pi_{55} &= (1/\Box^4)D^4B^4\bar{D}^4\bar{B}^4,
\end{aligned}
$$

where the obvious notations were introduced:

$$\Pi_{ab}\equiv\Pi_a\Pi_b, \quad \Pi_{ab}=\sum_{(\prime)}\Pi^{(\prime)}_{ab}. \qquad (3.2)$$

Note now that the SU(4) $\simeq$ SO(6) is well-known to be the maximal automorphism group (internal symmetry) of $N=4$ superalgebra,[19] whereas our constructions possess only the SU(2) $\times$ SU(2) $\simeq$ SO(4) internal symmetry. Consequently, the natural next step is to unite the $N=2\times 2$ irreducible superprojectors (3.1) into the ones with SU(4) internal symmetry. It can be done easily by using the following rules concerning the decomposition of SU(4)-irreps into the SU(2) $\times$ SU(2) ones:

$$
\begin{aligned}
4 &= (1,2)+(2,1),\\
6 &= 2(1,1)+(2,2),\\
10 &= (1,3)+(3,1)+(2,2), \qquad (3.3)\\
15 &= (1,1)+(1,3)+(3,1)+2(2,2),\\
20 &= 2(1,2)+2(2,1)+(2,3)+(3,2),\\
20' &= 3(1,1)+2(2,2)+(3,3).
\end{aligned}
$$

Application of these branching rules to superprojectors (3.1) gives rise to the standard SU(4)-invariant $N=4$ irreducible superprojectors[8] as follows:

$$
\begin{aligned}
\Pi_{\langle 0,0,1\rangle} &= (1/\Box^4)D^8\bar{D}^8=\Pi_{55},\\
\Pi_{\langle 8,0,1\rangle} &= (1/\Box^4)\bar{D}^8D^8=\Pi_{00},\\
\Pi_{\langle 1,1/2,4\rangle} &= (1/\Box^4)\bar{D}_{\dot\alpha A}D^8(\bar{D}^7)^{\dot\alpha A}=\Pi_{45}+\Pi_{54},\\
\Pi_{\langle 7,1/2,\bar{4}\rangle} &= (1/\Box^4)D^{\alpha A}\bar{D}^8(D^7)_{\alpha A}=\Pi_{10}+\Pi_{01},\\
\Pi_{\langle 2,0,10\rangle} &= (1/\Box^4)\bar{D}_{AB}D^8(\bar{D}^6)^{AB}=\Pi_{25}+\Pi_{52}+\Pi''_{44},\\
\Pi_{\langle 6,0,\overline{10}\rangle} &= (1/\Box^4)D^{AB}\bar{D}^8(D^6)_{AB}\\
&= \Pi_{20}+\Pi_{02}+\Pi''_{11},\\
\Pi_{\langle 2,1,6\rangle} &= (1/2\Box^4)\bar{D}_{\dot\alpha\dot\beta[AB]}D^8(\bar{D}^6)^{\dot\alpha\dot\beta[AB]}\\
&= \Pi_{35}+\Pi_{53}+\Pi'_{44},\\
\Pi_{\langle 6,1,\bar{6}\rangle} &= (1/2\Box^4)D^{[AB]}_{\alpha\beta}\bar{D}^8(D^6)^{\alpha\beta}_{[AB]}\\
&= \Pi_{30}+\Pi_{03}+\Pi'_{11}, \qquad (3.4)\\
\Pi_{\langle 3,3/2,\bar{4}\rangle} &= (1/\Box^4)(\bar{D}^3)^{A}_{\dot\alpha\dot\beta\dot\gamma}D^8(\bar{D}^5)^{\dot\alpha\dot\beta\dot\gamma}_{A}=\Pi'_{43}+\Pi'_{34},\\
\Pi_{\langle 5,3/2,4\rangle} &= (1/\Box^4)(D^3)^{\alpha\beta\gamma}_{A}\bar{D}^8(D^5)^{A}_{\alpha\beta\gamma}=\Pi'_{13}+\Pi'_{31},\\
\Pi_{\langle 3,1/2,20\rangle} &= -(1/6\Box^4)(\bar{D}^3)_{\dot\alpha[AB]C}D^8(\bar{D}^5)^{\dot\alpha[AB]C}\\
&= \Pi_{15}+\Pi_{51}+\Pi''_{34}+\Pi''_{43}+\Pi'_{42}+\Pi'_{24},\\
\Pi_{\langle 5,1/2,\overline{20}\rangle} &= -(1/6\Box^4)(D^3)^{\alpha[AB]C}\bar{D}^8(D^5)_{\alpha[AB]C}\\
&= \Pi_{40}+\Pi_{04}+\Pi''_{31}+\Pi''_{13}+\Pi'_{21}+\Pi'_{12},\\
\Pi_{\langle 4,2,1\rangle} &= (1/\Box^4)(D^4)^{\alpha\beta\gamma\delta}\bar{D}^8(D^4)_{\alpha\beta\gamma\delta}=\Pi'_{33},\\
\Pi_{\langle 4,1,15\rangle} &= (1/\Box^4)(D^4)^{\alpha\beta A}_{B}\bar{D}^8(D^4)^{B}_{\alpha\beta A}\\
&= \Pi'_{14}+\Pi'_{41}+\Pi_{23}+\Pi_{32}+\Pi''_{33},\\
\Pi_{\langle 4,0,20'\rangle} &= (1/\Box^4)(D^4)^{[AB]}_{[CD]}\bar{D}^8(D^4)^{[CD]}_{[AB]}\\
&= \Pi_{05}+\Pi_{50}+\Pi''_{12}+\Pi''_{21}+\Pi''_{24}+\Pi''_{42}\\
&\quad +\Pi''_{14}+\Pi''_{41}+\Pi_{22}+\Pi'''_{33}.
\end{aligned}
$$

We identify these $N=4$ irreps according to the standard classification[8–10] w.r.t. the eigenvalues of Casimir operators of the $N=4$ superalgebra: $N=4$ supercharge and superspin, and the dimension of SU(4) irreducible representation, respectively.

IV. $N=4$ IRREDUCIBLE SCALAR SUPERFIELDS

Now the construction of irreducible $N+4$ scalar superfields becomes elementary. Taking any irreducible $N=2$ scalar superfield (all these superfields have been explicitly constructed[16]), we consider each of its independent components as a certain irreducible scalar $N=2$ superfield, but now with possible SL(2,C) $\times$ SU(2) external indices on every component. This procedure immediately allows us to construct all 36 superfields, introduced in the previous section. Some of these $N=2\times 2$ superfields have to be further reduced according to the rules (3.1). Some work is necessary to resolve linear dependences among the components of the initial $N=2\times 2$ superfield in order to pick out the irreducible $N=2\times 2$ superfields, but the actual labor is much more simple than if one starts with a general $N=4$ superfield. Having obtained 50 irreducible $N=2\times 2$ superfields,

TABLE II. The space-time spin distribution of the independent components of ϕ_{33}.

Spin	4	$\frac{7}{2}$	3	$\frac{5}{2}$	2	$\frac{3}{2}$	1	$\frac{1}{2}$	0
Number of degrees of freedom	9	64	203	384	495	480	375	224	70

one can construct all 15 irreducible SU(4)-extended superfields elementary in accordance with Eq. (3.4). At this stage only the branching rules (3.3) are needed in order to arrange the products of anticommuting superspace coordinates in all orders in the decomposition of an embracing $N=4$ superfield. It automatically gives rise to the SU(4)-covariant components in terms of the $SU(2)\times SU(2)$ ones.

To be more explicit, consider the particular example of the superfield ϕ_{33}. This $N=2\times 2$ superfield has the structure of $\phi_{(1,0,2)}$ with respect to each $N=2$ supersymmetry it includes. The field component contents of this $N=2\times 2$ superfield ϕ_{33} is just Table I "squared." It performs the reducible supermultiplet with $(24+24)\times(24+24)=1152+1152$ components (for comparison, a general $N=4$ scalar superfield contains $65\,536+65\,536$ components). The space-time spin distribution of the independent components of ϕ_{33} is given by Table II. The superfield ϕ_{33} is reducible and decomposes into three irreducible $N=2\times 2$ superfields: ϕ'_{33} with the highest spin 4, ϕ''_{33} with the highest spin 3, and ϕ'''_{33} with the highest spin 2. Clearly, the last one is of special interest. To pick out ϕ'''_{33}, we must eliminate ϕ'_{33} and ϕ''_{33} from ϕ_{33}. One of the ways to do it is to require the vanishing of higher spin components in ϕ_{33} (with spin more than 2). Supersymmetry then also requires the vanishing of whole superfields ϕ'_{33} and ϕ''_{33}. The actual calculations, being elementary, turn out to be very tedious because of the great number of components. So, we present here only spin contents of ϕ'_{33} (640 + 640 components) (see Table III) and that of ϕ''_{33} (384 + 384 components) (see Table IV).

TABLE IV. The spin contents of ϕ''_{33}.

Spin	3	$\frac{5}{2}$	2	$\frac{3}{2}$	1	$\frac{1}{2}$	0
Number of degrees of freedom	7	48	140	224	210	112	27

The remaining $N=2\times 2$ irreducible superfield ϕ'''_{33} with the highest spin 2 consists of 128 + 128 independent components. Information about them is presented in Table V. The spin distribution for the components of the superfield ϕ'''_{33} is shown in Table VI.

In the remainder of this section, we present the result of the explicit calculation of the quadratic Lagrangian in terms of components of the superfield ϕ'''_{33}:

$$\begin{aligned}
L &= \int d^4\theta\, d^4\bar\theta\, d^4\eta\, d^4\bar\eta\, (\phi'''_{33})^2 \\
&= \tfrac{1}{4} TT^{(2)}_{\mu\nu} TT^{(2)}_{\mu\nu} + \tfrac{1}{2}(ZT^{(3/2)\alpha i,}_{\mu}\, i\, \partial_{\alpha\dot\alpha} \bar{Z}T^{(3/2)\dot\alpha}_{\mu i,} + \mathrm{LT}) - 4TT^{\alpha\beta}TT_{\alpha\beta} - 2(TI'^{,ij}_{\mu} \Box TI'_{\mu,ij} + \mathrm{LT}) \\
&\quad - (TG_{\dot\alpha\dot\beta}\, \partial^{\alpha\dot\alpha} \partial^{\beta\dot\beta} T\bar{G}_{\alpha\beta} + \mathrm{LT}) - 2ZZ^{\alpha\beta i,j} \partial_{\alpha\dot\alpha}\, \partial_{\beta\dot\beta} \overline{ZZ}^{\dot\alpha\dot\beta}_{i,j} - 4Z\bar{Z}'^{i,j}_{\alpha\dot\alpha}\, \partial^{\alpha\dot\beta}\, \partial^{\beta\dot\alpha} \bar{Z}Z'_{\dot\beta\beta i,j} + (TZ^{,i}_{\dot\alpha}\, i\, \partial^{\alpha\dot\alpha} T\bar{Z}_{\alpha,i} + \mathrm{LT}) \\
&\quad - (ZG^{i,}_{\dot\alpha}\, i\, \partial^{\alpha\dot\alpha} \Box \overline{ZG}_{\alpha i,} + \mathrm{LT}) + (Z\bar{G}^{\alpha i,}\, i\, \partial_{\alpha\dot\alpha} \bar{Z}G^{\dot\alpha}_{i,} + \mathrm{LT}) + (ZI^{i,jk\alpha}\, i\, \partial_{\alpha\dot\alpha} \Box \bar{Z}I^{\dot\alpha}_{i,jk} + \mathrm{LT}) \\
&\quad + TT\cdot TT + II^{ij,kl} \Box^2 II_{ij,kl} + (IG^{ij,} \Box I\bar{G}_{ij,} + \mathrm{LT}) + ZZ^{i,j} \Box \overline{ZZ}_{i,j} + Z\bar{Z}^{i,j} \bar{Z}Z_{i,j} + G\bar{G}\cdot\bar{G}G + GG\, \Box^2 \overline{GG},
\end{aligned} \tag{4.1}$$

where the abbreviation LT is used to designate the similar additional terms with letters transposed. Clearly, the Lagrangian (4.1) is, in fact, formal, since the higher spin (2, $\frac{3}{2}$, and 1) field components satisfy differential constraints in space-time. Alternatively, explicit introduction of the "pure spin" projectors would give rise to nonlocalities in Eq. (4.1). The way to overcome this difficulty is discussed in Sec. V. For the sake of brevity, the $N=2\times 2$ supersymmetry transformation laws for the components of ϕ'''_{33} are not presented here. They may be obtained easily from Eq. (2.22).

TABLE III. The spin contents of ϕ'_{33}.

Spin	4	$\frac{7}{2}$	3	$\frac{5}{2}$	2	$\frac{3}{2}$	1	$\frac{1}{2}$	0
Number of degrees of freedom	9	64	196	336	350	224	84	16	1

V. LINEARIZED $N=4$ CONFORMAL SUPERGRAVITY

To construct $N=4$ conformal supergravity theory, one usually follows the common practice[5–7,14] of gauging the rigid superconformal group of $N=4$ super-Yang–Mills (SYM) theory. In this approach the knowledge of the on-shell $N=4$ SYM theory[20] is sufficient. After imposing the conventional constraints on curvatures, the additional matter fields are needed to close the superalgebra. As a result, this way of construction turns out to be nonsystematic and somewhat involved.

The alternative superspace approach to $N=4$ conformal supergravity was developed by Howe.[11] In the context of this method the superspace constraints for the full nonlinear $N=4$ conformal supergravity theory were formulated and partially solved.[11] However, the superspace action was not found. It should be noted, by the way, that this action may be found, in principle, as a counterterm for an $N=4$ SYM interacting with an external $N=4$ conformal supergravity.[14]

A way to construct the linearized $N=4$ conformal su-

TABLE V. The field component contents of ϕ'''_{33}. The highest spin is 2. The highest $SU(2)\times SU(2)$ representation is (3,3). All components are taken to be irreducible w.r.t. $SL(2,C)\times SU(2)\times SU(2)$. The notation for components are by their origin in the superfield $\phi_{(1,0,2)}$, which was "squared" to get ϕ_{33}.

Components	Spin	SL(2,C) irrep	SU(2)×SU(2) irrep	Number of degrees of freedom
$TT^{(2)}_{\mu\nu}$	2	(1,1)	(1,1)	5
$ZT^{(3/2)i,}_{\mu\alpha}, \bar{Z}T^{(3/2)i,}_{\mu\dot{\alpha}}$	$\frac{3}{2}$	$(1,\frac{1}{2})+(\frac{1}{2},1)$	(2,1)	8 + 8
$TZ^{(3/2),i}_{\mu\alpha}, T\bar{Z}^{(3/2),i}_{\mu\dot{\alpha}}$	$\frac{3}{2}$	$(1,\frac{1}{2})+(\frac{1}{2},1)$	(1,2)	8 + 8
$TT_{\alpha\beta}$	1	(1,0)	(1,1)	3
$TI'^{\,,ij}_{\mu}$	1	$(\frac{1}{2},\frac{1}{2})$	(1,3)	9
$IT'^{ij,}_{\mu}$	1	$(\frac{1}{2},\frac{1}{2})$	(3,1)	9
$Z\bar{Z}'^{i,j}_{\mu}, \bar{Z}Z'^{i,j}_{\mu}$	1	$(\frac{1}{2},\frac{1}{2})$	(2,2)	12 + 12
$TG_{\dot{\alpha}\dot{\beta}}, T\bar{G}_{\alpha\beta}$	1	(0,1) + (1,0)	(1,1)	3 + 3
$GT_{\dot{\alpha}\dot{\beta}}, \bar{G}T_{\alpha\beta}$	1	(0,1) + (1,0)	(1,1)	3 + 3
$ZZ^{i,j}_{\alpha\beta}, \overline{ZZ}^{i,j}_{\dot{\alpha}\dot{\beta}}$	1	(1,0) + (0,1)	(2,2)	12 + 12
$ZG^{i,}_{\dot{\alpha}}, \overline{ZG}^{i,}_{\alpha}$	$\frac{1}{2}$	$(0,\frac{1}{2})+(\frac{1}{2},0)$	(2,1)	4 + 4
$GZ^{,i}_{\dot{\alpha}}, \overline{GZ}^{,i}_{\alpha}$	$\frac{1}{2}$	$(0,\frac{1}{2})+(\frac{1}{2},0)$	(1,2)	4 + 4
$TZ^{,i}_{\dot{\alpha}}, T\bar{Z}^{,i}_{\alpha}$	$\frac{1}{2}$	$(0,\frac{1}{2})+(\frac{1}{2},0)$	(1,2)	4 + 4
$ZT^{i,}_{\dot{\alpha}}, \bar{Z}T^{i,}_{\alpha}$	$\frac{1}{2}$	$(0,\frac{1}{2})+(\frac{1}{2},0)$	(2,1)	4 + 4
$ZI^{i,jk}_{\alpha}, \bar{Z}I^{i,jk}_{\dot{\alpha}}$	$\frac{1}{2}$	$(\frac{1}{2},0)+(0,\frac{1}{2})$	(2,3)	12 + 12
$IZ^{ij,k}_{\alpha}, I\bar{Z}^{ij,k}_{\dot{\alpha}}$	$\frac{1}{2}$	$(\frac{1}{2},0)+(0,\frac{1}{2})$	(3,2)	12 + 12
$Z\bar{G}^{i,}_{\alpha}, \bar{Z}G^{i,}_{\dot{\alpha}}$	$\frac{1}{2}$	$(\frac{1}{2},0)+(0,\frac{1}{2})$	(2,1)	4 + 4
$G\bar{Z}^{,i}_{\dot{\alpha}}, \bar{G}Z^{,i}_{\alpha}$	$\frac{1}{2}$	$(0,\frac{1}{2})+(\frac{1}{2},0)$	(1,2)	4 + 4
TT	0	(0,0)	(1,1)	1
$Z\bar{Z}^{i,j}, \bar{Z}Z^{i,j}$	0	(0,0)	(2,2)	4 + 4
$II^{ij,kl}$	0	(0,0)	(3,3)	9
$G\bar{G}, \bar{G}G$	0	(0,0)	(1,1)	1 + 1
$ZZ^{i,j}, \overline{ZZ}^{i,j}$	0	(0,0)	(2,2)	4 + 4
$IG^{ij}, I\bar{G}^{ij,}$	0	(0,0)	(3,1)	3 + 3
$GI^{,ij}, \bar{G}I^{,ij}$	0	(0,0)	(1,3)	3 + 3
$GG, \overline{GG}$	0	(0,0)	(1,1)	1 + 1

pergravity is to construct the multiplet of currents of $N=4$ superconformal symmetries. One may hope that the knowledge of the structure of $N=2\times 2$ and $N=4$ irreducible superfields can, in particular, be useful for these aims. Hence we are interested in the irreducible superfields with the minimal $(128+128)$ number of off-shell degrees of freedom. This number $128+128$ coincides with that of states in the smallest on-shell massive $N=4$ multiplet.[4]

Direct inspection reveals that three superfields—the real versions of ϕ_{05} (or ϕ_{50}), ϕ_{22}, and ϕ'''_{33}—contain the desired number of degrees of freedom (with the highest spin 2). All these irreducible superfields are the $SU(2)\times SU(2)$ extended ones, but the new feature of $N=4$ superspace in comparison with its $N=2$ predecessor is that any of these $(128+128)$ $N=2\times 2$ superfields with less than SU(4) internal symmetry by construction can, in fact, be rearranged into the superfields with SU(4) internal symmetry. This phenomenon was first found by Howe and Lindström[21] (see also the related work[22]), who have constructed the supercurrent for the dual version of the $N=4$ Maxwell theory, in which one of the scalars was replaced by an antisymmetric tensor.[23] Consequently, we suppose that the above-mentioned superfields with $128+128$ independent field components are just irreducible nonconformal supercurrents of dual versions of $N=4$ gauge theory, but, in general, with less than SU(4) internal symmetry. In particular, the SU(4)-rearranged irreducible superfield ϕ'''_{33} is just the $N=4$ conformal supercurrent.

In the remainder of this section we outline a simple systematic way to construct the linearized $N=4$ conformal supergravity theory by making use of the $N=4$ conformal supercurrent ϕ'''_{33} in the SU(4)-invariant form.

First of all, we change the dimensions of some components of ϕ'''_{33} to put their values in correspondence with the canonical ones according to their spin. Explicitly,

$$
\begin{aligned}
&TZ^{\alpha,i}_{\mu}\to(\partial^{\alpha\dot{\alpha}}/\Box)TZ^{,i}_{\dot{\alpha}\mu},\quad ZT^{\alpha i}_{\mu}\to(\partial^{\alpha\dot{\alpha}}/\Box)ZT^{i,}_{\dot{\alpha}\mu},\\
&TI'^{,ij}_{\mu}\to\Box TI'^{,ij}_{\mu},\quad IT'^{ij,}_{\mu}\to\Box IT'^{ij,}_{\mu},\\
&TT_{\alpha\beta}\to(1/2\Box)\partial_{(\alpha\dot{\alpha}}TT^{\dot{\alpha}}_{\beta)},\quad TT_{\alpha\dot{\alpha}}\equiv\tfrac{1}{2}(\sigma_{\mu})_{\alpha\dot{\alpha}}TT'_{\mu},\\
&Z\bar{Z}'^{i,j}_{\mu}\to\Box Z\bar{Z}'^{i,j}_{\mu},\quad ZT^{i,}_{\dot{\alpha}}\to\Box ZT^{i,}_{\dot{\alpha}},\quad TZ^{,i}_{\dot{\alpha}}\to\Box TZ^{,i}_{\dot{\alpha}},\\
&Z\bar{G}^{i,}_{\alpha}\to\Box Z\bar{G}^{i,}_{\alpha},\quad \bar{G}Z^{,i}_{\alpha}\to\Box\bar{G}Z^{,i}_{\alpha},\\
&TG_{\dot{\alpha}\dot{\beta}}\to\Box TG_{\dot{\alpha}\dot{\beta}},\quad ZZ^{i,j}_{\alpha\beta}\to\Box ZZ^{i,j}_{\alpha\beta},\\
&(\sigma_{\mu})^{\alpha\dot{\alpha}}TG_{\dot{\alpha}\dot{\beta}}\equiv TG^{\ \alpha}_{\mu\ \dot{\beta}}\equiv\tfrac{1}{2}(\sigma_{\nu})^{\alpha}_{\beta}TG_{\mu\nu},\\
&(\sigma_{\mu})^{\alpha\dot{\alpha}}ZZ^{i,j}_{\alpha\beta}\equiv ZZ^{i,j\,\dot{\alpha}}_{\mu\ \beta}\equiv\tfrac{1}{2}(\sigma_{\nu})^{\dot{\alpha}}_{\beta}ZZ^{i,j}_{\mu\nu},
\end{aligned}
\tag{5.1}
$$

TABLE VI. The spin distribution for the components of the superfield ϕ'''_{33}.

Spin	2	$\frac{3}{2}$	1	$\frac{1}{2}$	0
Number of degrees of freedom	5	32	81	96	42

$$TG_{\mu\nu}=\begin{pmatrix}(\sigma_{\mu\nu})_{\alpha\beta}TG^{\alpha\beta}\\(\tilde\sigma_{\mu\nu})_{\dot\alpha\dot\beta}\overline{TG}^{\dot\alpha\dot\beta}\end{pmatrix},\quad ZZ^{i,j}_{\mu\nu}=\begin{pmatrix}(\sigma_{\mu\nu})^{\alpha\beta}ZZ^{i,j}_{\alpha\beta}\\(\tilde\sigma_{\mu\nu})^{\dot\alpha\dot\beta}\overline{ZZ}^{i,j}_{\dot\alpha\dot\beta}\end{pmatrix},$$

$$ZI^{\alpha i,jk}\to(\partial^{\alpha\dot\alpha}/\Box^2)ZI^{i,jk}_{\dot\alpha},\quad GZ^{,i}_{\dot\alpha}\to\Box GZ^{,i}_{\dot\alpha},$$

$$IZ^{\alpha ij,k}\to(\partial^{\alpha\dot\alpha}/\Box^2)IZ^{ij,k}_{\dot\alpha},\quad ZG^{i,}_{\dot\alpha}\to\Box ZG^{i,}_{\dot\alpha},$$

$$TT\to\Box TT,\quad G\overline{G}\to\Box G\overline{G},\quad Z\overline{Z}^{i,j}\to\Box Z\overline{Z}^{i,j},$$

$$II^{ij,kl}\to\Box^2 II^{ij,kl},\quad ZZ^{i,j}\to\Box ZZ^{i,j},\quad GG\to\Box GG.$$

In terms of the redefined components (5.1), the superspace quadratic Lagrangian reads as follows:

$$\begin{aligned}L\to\int d^{16}\theta\,\phi'''_{33}\Box^2\phi'''_{33}
&=\tfrac14 TT^{(2)}_{\mu\nu}\Box^2TT^{(2)}_{\mu\nu}+\tfrac12(TZ^{(3/2),i}_{\mu\dot\alpha}i\,\partial^{\alpha\dot\alpha}\Box\overline{TZ}^{(3/2)}_{\mu\alpha,i}+LT)-2TT'_{\mu}\Box TT'_{\mu}-2(TI'^{,ij}_{\mu}\Box TI'_{\mu,ij}+LT)-2Z\overline{Z}'^{i,j}_{\mu}\Box\overline{Z}Z'_{\mu i,j}\\
&\quad+ZZ^{i,j}_{\mu\lambda}\,\partial_\mu\,\partial_\nu\overline{ZZ}_{\nu\lambda i,j}+(TG_{\mu\lambda}\,\partial_\mu\,\partial_\nu\,\overline{TG}_{\nu\lambda}+LT)+(TZ^{,i}_{\dot\alpha}i\,\partial^{\alpha\dot\alpha}\overline{TZ}_{\alpha,i}+LT)-(ZI^{i,jk}_{\dot\alpha}i\,\partial^{\alpha\dot\alpha}\overline{Z}I_{\alpha i,jk}+LT)\\
&\quad+(Z\overline{G}^{i,\alpha}i\,\partial_{\alpha\dot\alpha}\overline{Z}G^{\dot\alpha}_{i,}+LT)-(ZG^{i,}_{\dot\alpha}i\,\partial^{\alpha\dot\alpha}\Box\overline{ZG}_{\alpha i,}+LT)+TT\cdot TT+Z\overline{Z}^{i,j}\overline{Z}Z_{i,j}+G\overline{G}\cdot\overline{G}G+II^{ij,kl}II_{ij,kl}\\
&\quad+ZZ^{i,j}\Box\overline{ZZ}_{i,j}+(IG^{ij,}\Box\overline{IG}_{ij,}+LT)+GG\,\Box^2\overline{GG}.\end{aligned}\tag{5.2}$$

The dimensional redefinitions (5.1) give rise to some nonlocal terms in the $N=4$ supersymmetry transformation laws (which are not given here) for the redefined components. However, it is easy to show that, if one introduces gauge fields instead of the "pure spin" ones by means of corresponding spin projectors,

$$\begin{aligned}&TT^{(2)}_{\mu\nu}=P^{(2)}_{\mu\nu\lambda\rho}\,g_{\lambda\rho},\quad TZ^{(3/2)}_{\mu}=P^{(3/2)}_{\mu\nu}\psi_\nu,\\&X'_\mu=P^{\perp}_{\mu\nu}V_\nu,\end{aligned}\tag{5.3}$$

where X'_μ stands for TT'_μ, $TI'^{,ij}_{\mu}$, $IT'^{ij,}_{\mu}$, and $Z\overline{Z}'^{i,j}_{\mu}$, all the nonlocal terms in the supersymmetry transformation laws can be absorbed into gauge transformations:

$$\begin{aligned}&\delta g_{\mu\nu}=\partial_{(\mu}\xi_{\nu)}+\eta_{\mu\nu}\sigma,\\&\delta\psi^i=\partial_\mu\epsilon^i+\gamma_\mu\pi^i,\quad\delta V_\mu=\partial_\mu\varkappa.\end{aligned}\tag{5.4}$$

Clearly, it results in the $N=4$ supersymmetry transformation superalgebra with central charges, the latter being given by the gauge transformations (5.4).

Now we simply arrange the components into SU(4)-irreducible multiplets as follows:

$$\psi^A_\mu=\begin{pmatrix}\psi^{i,}_{\mu\alpha}\\\psi^{,i}_{\mu\alpha}\\\bar\psi^{i,\dot\alpha}_{\mu}\\\bar\psi^{,i\dot\alpha}_{\mu}\end{pmatrix},\quad V^A_{\mu B}=\begin{pmatrix}V^{ij,}_{\mu}\\V^{i,j}_{\mu}\\V_\mu\\V^{,ij}_{\mu}\end{pmatrix},\quad T^{[AB]}_{\mu\nu}=\begin{pmatrix}TG_{\mu\nu}\\ZZ^{i,j}_{\mu\nu}\\GT_{\mu\nu}\end{pmatrix},\quad\chi^{[AB]C}=\begin{bmatrix}\overline{Z}T^{i,}_{\alpha}\\Z\overline{G}^{i,}_{\alpha}\\\overline{Z}I^{i,jk}_{\alpha}\\I\overline{Z}^{ij,k}_{\alpha}\\\overline{G}Z^{,i}_{\alpha}\\T\overline{Z}^{,i}_{\alpha}\\ZT^{i,\dot\alpha}\\ZG^{i,\dot\alpha}\\ZI^{i,jk\dot\alpha}\\IZ^{ij,k\dot\alpha}\\GZ^{,i\dot\alpha}\\TZ^{,i\dot\alpha}\end{bmatrix},$$

$$\Lambda^A=\begin{pmatrix}\overline{ZG}^{i,}_{\alpha}\\\overline{GZ}^{,i}_{\alpha}\\ZG^{i,\dot\alpha}\\GZ^{,i\dot\alpha}\end{pmatrix},\quad D^{[AB]}_{[CD]}=\begin{pmatrix}G\overline{G}\\Z\overline{Z}^{i,j}\\II^{ij,kl}\\TT\end{pmatrix},\quad E^{AB}=\begin{pmatrix}IG^{ij,}\\ZZ^{i,j}\\GI^{,ij}\end{pmatrix},\quad\varphi=GG.\tag{5.5}$$

We also use spin projectors in their explicit form,

$$\begin{aligned}P^{(2)}_{\mu\nu\lambda\rho}&=\tfrac12P^{\perp}_{\mu(\lambda}P^{\perp}_{\rho)\nu}-\tfrac13P^{\perp}_{\mu\nu}P^{\perp}_{\lambda\rho},\\P^{(3/2)}_{\mu\nu}&=\eta_{\mu\nu}-\tfrac23(\partial_\mu\,\partial_\nu/\Box)-\tfrac13\gamma_\mu\gamma_\nu\\&\quad-\tfrac13(\hat\partial\gamma_{[\mu}\,\partial_{\nu]}/\Box),\\P^{\perp}_{\mu\nu}&=\eta_{\mu\nu}-\partial_\mu\partial_\nu/\Box,\end{aligned}\tag{5.6}$$

to rewrite the Lagrangian (5.2) in the SU(4)-invariant convenient form using the four-component notation

$$\begin{aligned}L&=(R^2_{\mu\nu}-\tfrac13R^2)_{\rm lin}+\epsilon_{\mu\nu\lambda\rho}(\overline{R_\mu-\tfrac13\gamma_\mu\hat R})_A i\gamma_5\gamma_\nu\partial_\lambda\\&\quad\times(R_\rho-\tfrac13\gamma_\rho\hat R)^A-F^2_{\mu\nu}(V)+T^{[AB]}_{\mu\lambda}\,\partial_\mu\,\partial_\nu\\&\quad\times\overline{T}_{\nu\lambda[AB]}+\bar\chi^{[AB]C}i\,\hat\partial\chi_{[AB]C}-\overline{\Lambda}^A i\,\hat\partial\,\Box\Lambda_A\\&\quad+D^{[AB]}_{[CD]}D^{[CD]}_{[AB]}+E^{AB}\Box\overline{E}_{AB}+\varphi\,\Box^2\varphi,\end{aligned}\tag{5.7}$$

where we have introduced the standard notation for the

Ricci tensor $R_{\mu\nu}$, the scalar curvature R, the gravitino field strength

$$R^{A}_{\mu} = \epsilon_{\mu\nu\lambda\sigma} i\gamma_5\gamma_\nu\, \partial_\lambda \psi^A_\sigma, \quad \hat{R}^{A} = \gamma_\mu R^{A}_{\mu}, \tag{5.8}$$

and the vector field strength

$$F_{\mu\nu}(V) = \partial_{[\mu} V_{\nu]}. \tag{5.9}$$

The components $V^A_{\mu B}, \chi^{[AB]C}$, and $D^{[AB]}_{[CD]}$, being the irreducible SU(4) tensors, satisfy the irreducibility conditions

$$V^A_{\mu A} = \epsilon_{ABCD}\chi^{[AB]C} = D^{[AB]}_{[AC]} = 0. \tag{5.10}$$

The theory (5.7) is just the well-known linearized Lagrangian of $N=4$ conformal supergravity theory. We stop our discussion here since the detailed component description of the linearized $N=4$ conformal supergravity theory itself has already been given in the literature.[5–7,14]

VI. CONCLUSIONS

In summary, we note that the methods we follow are, in fact, very general. They can be applied to the construction of irreducible superfields in any N-extended superspace in the four-dimensional space-time. Moreover, this approach can easily be generalized to higher dimensions of space-time, in particular, to $d=10$, in order to construct the irreducible superfields in $N=1$ or $N=2$ ten-dimensional superspace (see the related works[24]). We hope that the knowledge of the irreducible superfields in extended superspace may be useful in the analysis of extended supergravity theories.

ACKNOWLEDGMENTS

The authors would like to thank I. V. Tyutin and A. A. Tseytlin for useful discussions.

1. A. Salam and J. Strathdee, Nucl. Phys. B **76**, 477 (1974); Phys. Rev. D **11**, 1521 (1975); Fortschr. Phys. **26**, 57 (1978).
2. A. Galperin, E. Ivanov, S. Kalitzin, V. Ogievetsky, and E. Sokatchev, Class. Quantum Grav. **1**, 469 (1984); **2**, 155, (1985); A. Galperin, E. Ivanov, V. Ogievetsky, and E. Sokatchev, *ibid.* **2**, 601, 617 (1985); Commun. Math. Phys. **103**, 515 (1986).
3. E. Ahmed, S. Bedding, C. T. Card, M. Dumbrell, M. Nouri-Moghadam, and J. G. Taylor, J. Phys. A: Math. Gen. **18**, 2095 (1985).
4. B. de Wit and S. Ferrara, Phys. Lett. B **81**, 317 (1979).
5. *Superspace and Supergravity, Proceedings of the Nuffield Supergravity Workshop*, edited by S. W. Hawking and M. Roček (Cambridge U.P., Cambridge, 1981).
6. E. Bergshoeff, M. de Roo, and B. de Wit, Nucl. Phys. B **182**, 173 (1981); in Ref. 5, p. 237.
7. B. de Wit, in *Supergravity '81*, edited by S. Ferrara and J. G. Taylor (Cambridge U.P., Cambridge, 1982), p. 267.
8. W. Siegel and S. J. Gates, Jr., Nucl. Phys. B **189**, 295 (1981).
9. P. S. Howe, K. S. Stelle, and P. K. Townsend, Nucl. Phys. B **191**, 445 (1981); B **192**, 332 (1981).
10. V. Rittenberg and E. Sokatchev, Nucl. Phys. B **193**, 471 (1981).
11. P. S. Howe, Phys. Lett. B **100**, 389 (1981); Nucl. Phys. B **199**, 309 (1982).
12. V. Rivelles and J. G. Taylor, J. Phys. A: Math. Gen. **15**, 163 (1982).
13. E. Ivanov, S. Kalitzin, Nguyen Ai Viet, and V. Ogievetsky, J. Phys. A: Math. Gen. **18**, 3433 (1985).
14. E. S. Fradkin and A. A. Tseytlin, Phys. Rep. **119**, 233 (1985).
15. M. F. Sohnius, Phys. Rep. **128**, 39 (1985).
16. S. V. Ketov and I. V. Tyutin, Theor. Math. Fiz. **61**, 254 (1984); Pis'ma Zh. Eksp. Teor. Fiz. **39**, 573 (1984); Yad. Fiz. **41**, 1350 (1985); S. V. Ketov, Izv. Vyssh. Uchebn. Zaved. Fiz. No. 5, 93 (1986); No. 10, 31 (1986).
17. S. V. Ketov, B. B. Lokhvitsky, Ya. S. Prager, and I. V. Tyutin, TF SD AS USSR Reports Nos. 14, 15, and 16, 1985 (unpublished).
18. V. I. Ogievetsky and L. Mezincescu, Usp. Fiz. Nauk. **117**, 637 (1975); P. Fayet and S. Farrara, Phys. Rep. **32**, 249 (1977).
19. R. Haag, J. T. Łopuszanski, and M. F. Sohnius, Nucl. Phys. B **88**, 257 (1975).
20. L. Brink, J. H. Schwarz, and J. Scherk, Nucl. Phys. B **121**, 77 (1977); M. F. Sohnius and P. C. West, Phys. Lett. B **100**, 245 (1981).
21. P. S. Howe and U. Lindström, Nucl. Phys. B **181**, 487 (1981).
22. K. S. Stelle, LPTENS report 81/24, 1981 (unpublished).
23. M. F. Sohnius, K. S. Stelle, and P. C. West, Phys. Lett. B **92**, 123 (1980).
24. E. Bergshoeff, M. de Roo, and B. de Wit, Nucl. Phys. B **217**, 489 (1983); M. Villasante, Lett. Math. Phys. **11**, 351 (1986); R. Finkelstein and M. Villasante, J. Math. Phys. **27**, 1595 (1986).

Extended Weinberg–Witten theorem

Jan Łopuszański[a)]
University of Alberta, Edmonton, Alberta T6G 2Y1, Canada

(Received 16 July 1987; accepted for publication 16 December 1987)

A new method of proving the Weinberg–Witten theorem is presented. It is shown that this theorem can be extended to currents that, being carriers of a charge, cause simultaneously a flip of the helicity of the massless one-particle state.

I. INTRODUCTION

This paper is concerned with a theorem stated by Weinberg and Witten[1,2] in 1980. A similar conjecture as that presented in this theorem was published as early as 1962 in the pioneer work of Case and Gasiorowicz.[3] Weinberg and Witten assert that in a local field theory a massless particle of helicity h for

$$|h|>\tfrac{1}{2}(k+l) \qquad (1)$$

cannot be a carrier of a Poincaré covariant charge induced by a covariant local Noetherian current $\Phi^{(k,l)}$ where $k,l = 0,\frac{1}{2},1,\ldots$ indicate the transformation character of this current with respect to the Lorentz group.

The proof of this theorem[2] relies heavily upon relativistic covariance, locality, and positive definiteness of the metric in the Hilbert space, as well as upon the existence of the one-particle states. The method of proof consists of examining the matrix elements of a current sandwiched between two massless one-particle states. It is well known that the vanishing of these matrix elements implies the vanishing of the current viewed as an operator valued distribution in the Hilbert space. Let us denote any massless one-particle state characterized by momentum

$$p\equiv(p_0,p_1,p_2,p_3),\quad p^2=p^\mu p_\mu=0,\quad p_0\geqslant 0,$$

and helicity

$$h=0,\pm\tfrac{1}{2},\pm 1,\ldots$$

by $\psi(p,h)$ and a translationally covariant quantum field by $\Phi_N^{(k,l)}(x)$. The Lorentz transformation properties of the massless one-particle state are characterized by the indices (k,l) and N labels the components,[4] viz.,

$$U(a,A)\Phi_N^{(k,l)}U(a,A)^+=\sum_M (S^{-1})_N^M\Phi_M^{(k,l)}(\Lambda x+a).$$

Here $U(a,A)$ is a unitary operator in the Hilbert space representing the element of the Poincaré group characterized by

$$A\in \mathrm{SL}(2,C),$$

i.e.,

$$A=\begin{pmatrix} a & b\\ c & d\end{pmatrix},\quad ad-bc=1,$$

a,b,c,d—complex numbers,

and

$$a\equiv(a_0,a_1,a_2,a_3),\quad a_\mu=\bar a_\mu,\quad \mu=0,1,2,3,$$

$$S_N{}^M=S(A)_N{}^M$$

represents the same group element A in a space where the number of dimensions is determined by the tensor character of the field. Finally,

$$\Lambda_\mu{}^\nu=\Lambda(A)_\mu{}^\nu$$

is the four-dimensional Lorentz transformation corresponding to A. Using the notation

$$(\psi(p,h),\Phi_N^{(k,l)}(x))\psi'(p',h')\equiv\Phi_N^{(k,l)}(p,p';h,h')e^{i(p-p')x}$$

it can be shown[1,2] that

$$\Phi_N^{(k,l)}(p,p';h,h')=0,$$

for the scalar field $\Phi^{(0,0)}(x)$, $|h+h'|=0$,

for a spinor field $\Phi^{(1/2,0)}$ or $\Phi^{(0,1/2)}$, $|h+h'|\neq\frac{1}{2}$, (2)

for a vector field $\Phi^{(1/2,1/2)}$, $|h+h'|\neq 0$ or 1,

for a second rank tensor field $\Phi^{(1,1)}$, $|h+h'|\neq 0$, 1, or 2.

The Weinberg–Witten theorem as stated above concerns the particular case $h=h'$. If (1) is satisfied it follows from (2) that the current has to vanish.

Notice, however, that (2) does not exclude the existence of nonvanishing local Lorentz covariant currents which—being the carriers of a charge—obligatorily change the sign of the helicity when acting upon a one-particle state.[2,5] This possibility was discussed briefly in Ref. 2. It was pointed out that this effect does not contradict the Coleman–Mandula theorem[6] since (i) the change of a sign of helicity is related to a discrete symmetry (space reflection) to which the Coleman–Mandula theorem does not apply, (ii) the Coleman–Mandula theorem was proved only in the case of massive particles, and (iii) to preserve locality of the fields we always have to combine terms of opposite helicities.

A simple example of a vector current in a theory of a complex, free, (local) massless antisymmetric tensor field $F_{\mu\nu}$, $\mu,\nu=0,1,2,3$, which is neither self-dual or anti-self-dual, was briefly discussed in Ref. 2.

This observation seemed promising as far as the construction of an energy momentum tensor for a theory of a massless free field of helicity $|h|=2$ was concerned.

Unfortunately, as we shall see, the contribution coming from the matrix element

$$\Phi^{(k,l)}(p,p',h,h')\exp\{i(p-p')x\},$$

for $h'=-h$, $|h|\neq 0$, when integrated over $\mathbf{x}$, vanishes for

[a)] On leave of absence from the Institute of Theoretical Physics, University of Wroclaw, Wroclaw, Poland.

 0022-2488/88/051253-05$02.50

$\mathbf{p}' \to \mathbf{p}$; consequently the charge operator obtained from the current

$$Q_N = \int \Phi_{0\tilde{N}}(x) d^3x$$

vanishes when applied in the Hilbert space spanned by one-particle states. Here $\tilde{N}$ stands for a tensor index and is related to N by

$$N = (\mu,\tilde{N}), \quad \mu = 0,1,2,3.$$

If this is so in the framework of free field theory, we may expect that the same will be true *a fortiori* in a model of interacting fields conforming with the standard relativistically covariant, local theory in a separable Hilbert space of positive definite metric.

Thus our assertion is that the Weinberg–Witten theorem can be extended also to such helicity flipping currents.

II. A MISLEADING EXAMPLE

Consider a complex, free, (local) massless skew symmetric tensor field

$$F_{\mu\nu}, \quad \mu,\nu = 0,1,2,3,$$

which is neither self-dual nor anti-self-dual. This field is composed of local self-dual and anti-self-dual fields, viz.,

$$S_{\mu\nu} = u^{(1)}_{\mu\nu}(h = -1) + [u^{(2)}_{\mu\nu}(h = 1)]^+, \tag{3a}$$

$$A_{\mu\nu} = u^{(3)}_{\mu\nu}(h = 1) + [u^{(4)}_{\mu\nu}(h = -1)]^+, \tag{3b}$$

where $u^{(l)}_{\mu\nu}(x)$, $l = 1,2,3,4$, consists of creation operators only, prescribed to helicity h. Thus

$$F_{\mu\nu}(x) = S_{\mu\nu}(x) + A_{\mu\nu}(x)$$

and

$$u^{(1)} \neq u^{(4)}, \quad u^{(2)} \neq u^{(3)}.$$

Consider a local current

$$\Phi_\lambda(x) = i\{F^{\mu\nu}(\partial_\lambda F^+_{\mu\nu}) - (\partial_\lambda F^{\mu\nu})F^+_{\mu\nu}\}\colon(x)$$
$$\equiv i\colon\{F(\partial_\lambda F^+) - (\partial_\lambda F)F^+\}\colon(x).$$

Notice that

$$FF^+ = \colon(SA^+ + AS^+)\colon \tag{4}$$

as

$$SS^+ = AA^+ = 0.$$

The latter is the consequence of the product of self-dual and anti-self-dual antisymmetric tensor fields vanishing.[7]

Now it becomes clear why we need $F_{\mu\nu}$, which is neither self-dual nor anti-self-dual; otherwise $F(\partial_\lambda F)^+$ and $(\partial_\lambda F)F^+$ would vanish separately.

Let us denote the vacuum state by Ω: then $u^{(2)}\Omega$ and $u^{(3)}\Omega$ represent the one-particle state of helicity $h = +1$ and $u^{(1)}\Omega$ and $u^{(4)}\Omega$ of $h = -1$. The only contribution to the matrix elements of the current between one-particle states comes from the terms

$$u^{(1)}u^{(3)+}, \quad u^{(4)}u^{(2)+}, \quad u^{(3)}u^{(1)+}, \quad u^{(2)}u^{(4)+} \tag{5}$$

in (4); the action of these terms upon the one-particle states of fixed helicity causes the change of sign of the helicity of the particle.

This example seems to indicate how a nontrivial local Lorentz covariant vector current Φ_λ can be constructed for

$$|h| = |h'| = 1$$

—at least for the case of a free field—for which the Weinberg–Witten condition

$$h + h' = 0$$

in (2) is satisfied. This could be then extended to more complicated cases, in particular to construct the energy momentum tensor for a $|h| = 2$ field.

Unfortunately, a closer look into the structure of the massless tensor field brings disillusion. To see this let us consider

$$\colon F(\partial_\lambda F)^+\colon(x) \equiv \tfrac{1}{2}\colon S_{AB}(\partial_\lambda A^{\dot{A}\dot{B}})^+\colon + \tfrac{1}{2}\colon A_{\dot{A}\dot{B}}(\partial_\lambda S^{AB})^+\colon$$
$$= \sum_{m=-1}^{+1} (-1)^{m+1}\colon(S_m\,\partial_\lambda A^+_{-m} + A_m\,\partial_\lambda S^+_{-m})\colon(x), \tag{6}$$

where

$$F_{\mu\nu} \equiv \tfrac{1}{4}(\sigma_\mu)_{A\dot{C}}(\sigma_\nu)_{B\dot{D}}(s^{AB}\epsilon^{\dot{C}\dot{D}} + A^{\dot{C}\dot{D}}\epsilon^{AB}).$$

The indices A and $\dot{B}$ take the values 1 and 2:

$$\sigma^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

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

$$S_{-1} \equiv (1/\sqrt{2})S_{11} = (1/\sqrt{2})S^{22},$$
$$S_0 \equiv S_{12} = S_{21} = -S^{21},$$
$$S_1 \equiv (1/\sqrt{2})S_{22} = (1/\sqrt{2})S^{11},$$

and a similar relation for A_m, as well as

$$S_m \equiv u^{(1)}_m(h = -1) + u^{(2)+}_m(h = 1)$$
$$= \left(\frac{1}{2\pi}\right)^{3/2} \int \frac{d^3p}{2|\mathbf{p}|} D^{(1,0)}([p])_{m,-1}$$
$$\times\{a^{(1)+}([p];-1)e^{ipx} + a^{(2)}([p];1)e^{-ipx}\},$$

$$A_m = u^{(3)}_m(h = 1) + u^{(4)}_m(h = -1)$$
$$= \left(\frac{1}{2\pi}\right)^{3/2} \int \frac{d^3p}{2|\mathbf{p}|} D^{(0,1)}([p])_{m,-1}$$
$$\times\{a^{(3)+}([p];1)e^{ipx} + a^{(4)}([p];-1)e^{-ipx}\}.$$

Here $a^{(l)}([p];h)$, $l = 1,2,3,4$, stands for an annihilation operator for momentum p and helicity h. Also,

$$D^{(j,0)} = \overline{D^{(0,j)}}$$

is a $(2j+1)$-dimensional representation of the Lorentz group and $[p] \in SL(2,c)$ is the two-dimensional boost matrix, viz.,

$$\begin{pmatrix} a(p) & b(p) \\ c(p) & d(p) \end{pmatrix} = \frac{1}{\sqrt{2r(p_0 + p_3)}}$$
$$\times \begin{pmatrix} p_0 + p_3, & p_1 - ip_2 \\ p_1 + ip_2, & p_0 - p_3 + 2r \end{pmatrix}, \quad r > 0. \tag{7}$$

The expressions (5) appearing in (6) all have a similar structure, e.g., for $u^{(1)}u^{(3)+}$ we obtain

$$\iint \frac{d^3p}{2|\mathbf{p}|}\frac{d^3p^1}{2|\mathbf{p}'|}(a(p)c(p') - a(p')c(p))^2 p'_\lambda \times (a^{(1)}([p]; -1))^+ a^{(3)}([p'];1)e^{i(p-p')x}$$

(for details of the calculation see Sec. III). Upon integration over $\mathbf{x}$ of the matrix elements of this expression sandwiched between one-particle states, the term

$$\int d^3x\, e^{-i(\mathbf{p}-\mathbf{p}')\mathbf{x}}$$

approaches

$$\delta(\mathbf{p}-\mathbf{p}')$$

and consequently the contribution to the charge due to this matrix element vanishes as

$$\lim_{\mathbf{p}\to\mathbf{p}'}(a(p)c(p') - a(p')c(p)) = 0.$$

III. GENERAL RESULT

Let us investigate the general case of a local free massless field. It is well known that the field

$$f_{A_1\cdots A_r\dot B_1\cdots \dot B_S}, \quad A_l = 1,2,$$
$$\dot B_m = 1,2, \quad l = 1,\ldots,r, \quad m = 1,\ldots,s,$$

for $r = s + 2j > s$, has the form

$$\prod_{l=1}^{s} \partial_{A_l\dot B_l} S_{C_1\cdots C_{2j}(x)}, \tag{8}$$

where

$$S_{C_1\cdots C_{2j}} \equiv S_{\underline{C}}$$

is a local field symmetric in the spinor indices; (8) results as a consequence of the requirement of positive definiteness of the metric in the Hilbert space, which entails

$$\partial^{A_1\dot C} f_{A_1\cdots A_r\dot B_1\cdots \dot B_s} = 0, \tag{9a}$$

$$\partial^{C\dot B_1} f_{A_1\cdots A_r\dot B_1\cdots \dot B_s} = 0. \tag{9b}$$

For $s = r + 2j$,

$$f_{A_1\cdots A_r\dot B_1\cdots \dot B_s} = \prod_{l=1}^{r} \partial_{A_l\dot B_l} A_{\dot C_1\cdots \dot C_{2j}}, \tag{10}$$

where

$$A_{\dot B_1\cdots \dot B_{2j}} \equiv A_{\underline{\dot B}}$$

is again symmetric in the indices. Notice that $S_{\underline{C}}$ (as well as $A_{\underline{\dot C}}$) is a sum of two terms of helicity of opposite sign $|h| = j$; each of these terms comprises either creation or annihilation operators only. There can, but does not need to be any relation between $S_{\underline{C}}$ and $(A_{\underline{\dot C}})^+$ whatsoever [see, e.g., Eq. (3)]. The only nonvanishing contribution of the matrix element of a current taken between one-particle states can result from the sesquilinear form in the fields (8) and (10).

Let us start with the simplest case of a vector current

$$\Phi^{(1/2,1/2)}_{A\dot B}.$$

The most general expressions for the terms of which vector currents are constructed are

$$\lim_{y\to x} :\{T_A{}^{C_1\cdots C_{2j}}{}_{\dot B}{}^{\dot D_1\cdots \dot D_{2j}} S_{C_1\cdots C_{2j}}(x) A_{\dot D_1\cdots \dot D_{2j}}(y)\}:, \tag{11}$$

$$\lim_{y\to x} :\{T_A{}^{\underline{CD}}{}_{\dot B} S_{\underline{C}}(x) S_{\underline{D}}(y)\}:, \tag{12a}$$

$$\lim_{y\to x} :\{T_{A\dot B}{}^{\underline{\dot C\dot D}} A_{\underline{\dot C}}(x) A_{\underline{\dot D}}(y)\}:, \tag{12b}$$

where T is a covariant expression composed out of entries ϵ_{EF} and $\epsilon_{\dot E\dot F}$ of

$$\epsilon = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

$$\partial^x_{E\dot F} \equiv (\sigma^\mu)_{E\dot F}\,\partial^x_\mu \ \text{and}\ \partial^y_{E\dot F}.$$

Because of the symmetry properties of $S_{\underline{C}}$ and $A_{\underline{\dot D}}$ the quantity T in (11) can have only the following structure:

$$\epsilon_A{}^{C_1}\epsilon_{\dot B}{}^{\dot D_1}\prod_{l=2}^{2j}\partial^{C_l\dot D_l}, \tag{13a}$$

$$\partial_A{}^{\dot D_1}\partial^{C_1}{}_{\dot B}\prod_{l=2}^{2j}\partial^{C_l\dot D_l}, \tag{13b}$$

$$\partial_{A\dot B}\prod_{l=1}^{2j}\partial^{C_l\dot D_l}. \tag{14}$$

For $j > \frac{1}{2}$ (13) applied to $S_{\underline{C}}(x)A_{\underline{\dot D}}(y)$ yields a vanishing result as either

$$\partial^{C_l\dot D_l} S_{\underline{C}}(x) = 0$$

or

$$\partial^{C_l\dot D_l} A_{\underline{\dot D}}(y) = 0,$$

by virtue of (9). The same is true for (14) and $j > 0$.

Notice that (11) does not flip the helicity when constrained to the one-particle Hilbert space, while (12) does. Since the helicity preserving sesquilinear form (11) of the current vanishes for $j > \frac{1}{2}$, which coincides with (1), we proved in passing the Weinberg–Witten theorem for the case of a vector current.

The more interesting case, however, is that of expressions (12), which flip the helicity. We shall examine only the case (12a) as a similar reasoning can be applied in case (12b) *mutatis mutandis.* In this case the only nonvanishing contributions can arise from

$$\epsilon_{A_1}{}^{C_1}\partial^{D_1}{}_{\dot B} S_{C_1\underline{C}} S_{D_1}{}^{\underline{C}} = \partial^{D_1}{}_{\dot B}(S_{A\underline{C}} S_{D_1}{}^{\underline{C}}), \tag{15}$$

with

$$S_{A\underline{C}} \equiv S_{AC_2\cdots C_{2j}}$$

and

$$\partial_{A\dot B} S_{\underline{C}} S^{\underline{C}}. \tag{16}$$

In (15) the sum runs over the $(2j-1)$ and in (16) over the $2j$ indices. These expressions do not need to vanish.

Before we look closer at the problem of whether expressions (15) or (16) can yield a nonvanishing contribution to the charge let us still examine the case of a tensor current

$$\Phi^{(1,1)}_{A_1A_2\dot B_1\dot B_2}.$$

The most general expression for a sesquilinear form reads in this case as

$$\lim_{y\to x} :\{T_{A_1A_2}{}^{\underline{C}}{}_{\dot B_1\dot B_2}{}^{\underline{\dot D}} S_{\underline{C}}(x) A_{\underline{\dot D}}(y)\}:, \tag{17}$$

$$\lim_{y\to x} :\{T_{A_1A_2}{}^{\underline{C}\underline{D}}{}_{\dot B_1\dot B_2} S_{\underline{C}}(x) S_{\underline{D}}(y)\}:, \tag{18}$$

and a similar expression for $A_{\underline{C}}$.

In case (17) the structure of T is as follows:

$$\epsilon_{A_1}{}^{C_1}\epsilon_{A_2}{}^{C_2}\epsilon_{\dot B_1}{}^{\dot D_1}\epsilon_{\dot B_2}{}^{\dot D_2} \prod_{l=3}^{2j} \partial^{C_l\dot D_l}, \tag{19a}$$

$$\epsilon_{A_1}{}^{C_1}\epsilon_{\dot B_1}{}^{\dot D_1}\partial_{A_2}{}^{\dot D_2}\partial^{C_2}{}_{\dot B_2} \prod_{l=3}^{2j} \partial^{C_l\dot D_l}, \tag{19b}$$

$$\partial_{A_1}{}^{\dot D_1}\partial_{A_2}{}^{\dot D_2}\partial^{C_1}{}_{\dot B_1}\partial^{C_2}{}_{\dot B_2} \prod_{l=3}^{2j} \partial^{C_l\dot D_l}, \tag{19c}$$

$$\epsilon_{A_1A_2}\epsilon_{\dot B_1}{}^{\dot D_1}\partial^{C_1}{}_{\dot B_2} \prod_{l=2}^{2j} \partial^{C_l\dot D_l}, \tag{20a}$$

$$\epsilon_{A_1}{}^{C_1}\epsilon_{\dot B_1\dot B_2}\partial_{A_2}{}^{\dot D_1} \prod_{l=2}^{2j} \partial^{C_l\dot D_l}, \tag{20b}$$

$$\epsilon_{A_1}{}^{C_1}\epsilon_{\dot B_1}{}^{\dot D_1}\partial_{A_2\dot B_2} \prod_{l=2}^{2j} \partial^{C_l\dot D_l}, \tag{20c}$$

$$\epsilon_{A_1A_2}\epsilon_{\dot B_1\dot B_2} \prod_{l=1}^{2j} \partial^{C_l\dot D_l}, \tag{21a}$$

$$\partial_{A_1\dot B_1}\partial_{A_2\dot B_2} \prod_{l=1}^{2j} \partial^{C_l\dot D_l}. \tag{21b}$$

Expressions (19)–(21) yield a vanishing contribution for $j>1$ if applied to $S_{\underline{C}}A_{\underline{\dot D}}$, for similar reasons as in the case of the vector current. As the helicity is preserved here and (1) is satisfied this is the proof of the Weinberg–Witten theorem for the second rank tensor current.

Going over to the case when the helicity gets inverted, i.e., to case (18), we have the following possibilities:

$$\begin{aligned}&\epsilon_{A_1}{}^{C_1}\epsilon_{A_2}{}^{C_2}\partial^{D_1}{}_{\dot B_1}\partial^{D_2}{}_{\dot B_2}(S_{C_1C_2\underline{C}}S_{D_1D_2}{}^{\underline{C}})\\ &\quad=\partial^{D_1}{}_{\dot B_1}\partial^{D_2}{}_{\dot B_2}(S_{A_1A_2\underline{C}}S_{D_1D_2}{}^{\underline{C}}),\end{aligned} \tag{22a}$$

$$\partial^{C}{}_{\dot B_1}\partial^{D}{}_{\dot B_2}[S_{A_1C\underline{C}}S_{A_2D}{}^{\underline{C}}], \tag{22b}$$

$$\epsilon_{A_1A_2}\partial^{C}{}_{\dot B_1}\partial^{D}{}_{\dot B_2}S_{C\underline{C}}S_{D}{}^{\underline{C}}, \tag{23a}$$

$$\epsilon_{\dot B_1\dot B_2}S_{A_1\underline{C}}S_{A_2}{}^{\underline{C}}, \tag{23b}$$

$$\partial_{A_2\dot B_1}\partial^{D}{}_{\dot B_2}(S_{A_1\underline{C}}S_{D}{}^{\underline{C}}), \tag{23c}$$

$$\epsilon_{A_1A_2}\epsilon_{\dot B_1\dot B_2}S_{\underline{C}}S^{\underline{C}}, \tag{24a}$$

$$\partial_{A_1\dot B_1}\partial_{A_2\dot B_2}S_{\underline{C}}S^{\underline{C}}. \tag{24b}$$

In (22) the sum runs over the $(2j-2)$, in (23) over the $(2j-1)$, and in (24) over the $-2j$ indices. These expressions do not need to vanish.

Let us now take the next step and evaluate the contribution to the charge originating from the helicity flipping expressions.

Let us begin with

$$S_{\underline{C}}S^{\underline{C}} = S_{C_1\cdots C_{2j}}S^{C_1\cdots C_{2j}} = (2j)!\sum_{m=-j}^{j}(-1)^{j+m}S_mS_{-m}, \tag{25}$$

where

$$S_{\underbrace{1\cdots1}_{j-m}\underbrace{2\cdots2}_{j+m}} \equiv [(j-m)!(j+m)!]^{1/2}S_m, \quad m=-j,\ldots+j,$$

$$S^{\overbrace{1\cdots1}^{j-m}\overbrace{2\cdots2}^{j+m}} \equiv (-1)^{j+m}S_{-m}[(j-m)!(j+m)!]^{1/2}.$$

The free local field for $|h|=j$ reads as

$$\begin{aligned}S_m(x) &= \left(\frac{1}{2\pi}\right)^{3/2}\int\frac{d^3p}{2|\mathbf{p}|}D^{(j,0)}([p])_{m,-j}\\ &\quad\times\{a^+([p];-j)e^{ipx}+a([p];j)e^{-ipx}\},\\ m &= -j,\ldots+j.\end{aligned} \tag{26}$$

We have

$$\begin{aligned}&D^{(j,0)}([p])_{m,-j}\\ &\quad=[(2j)!/(j-m)!(j+m)!]^{1/2}a(p)^{j-m}c(p)^{j+m},\end{aligned} \tag{27}$$

where $a(p)$ and $c(p)$ are the entries of the SL(2,C) boost matrix (7).

If we insert (26) into (25) the subintegral expression involves

$$\begin{aligned}&\sum_{m=-j}^{+j}(-1)^{j+m}D^{(j,0)}([p])_{m,-j}D^{(j,0)}([p'])_{-m,-j}\\ &\quad=(a(p)c(p')-a(p')c(p))^{2j},\end{aligned}$$

where we used (27). Hence

$$\begin{aligned}S_{\underline{C}}S^{\underline{C}} &= 2j!\left(\frac{1}{2\pi}\right)^3\\ &\quad\times\iint\frac{d^3p}{2|\mathbf{p}|}\frac{d^3p'}{2|\mathbf{p}'|}(a(p)c(p')-a(p')c(p))^{2j}\\ &\quad\times:\{a^+([p],-j)e^{ipx}+a([p],h)e^{-ipx}\}\\ &\quad\times\{a^+([p'],-j)e^{ipx}+a([p]',j)e^{-ip'x}\}:.\end{aligned} \tag{28}$$

Notice, however, that expression (28), when confined to the one-particle Hilbert space, yields a vanishing contribution to the charge when integrated over $\mathbf{x}$ for $j>0$ for the same reason as in the example of Sec. II.

To make clear how this procedure can be generalized to cases when the summation runs over less than $2j$ indices $\underline{c}$, as, e.g., in (22) or (23), let us still consider the special case

$$\begin{aligned}S_{1\underline{C}}S_1{}^{\underline{C}} &= S_{1C_2\cdots C_{2j}}S_1{}^{C_2\cdots C_{2j}}\\ &= \sum_{m=-j}^{j}(-1)^{j+m}\frac{(2j-1)!}{(j-m-1)!(j+m)!}\\ &\quad\times[(j-m)!(j+m)!(j+m+1)!\\ &\quad\times(j-m-1)!]^{1/2}S_mS_{-(m+1)}.\end{aligned}$$

If we take into account (26) and (27) under the integral over $\mathbf{p}$ and $\mathbf{p}'$ we obtain the following expression:

$$\begin{aligned}&(2j-1)!\sum_{m=-j}^{j}(-1)^{j+m}\left[\frac{(j+m)!(j+m+1)!}{(j-m)!(j-m-1)!}\right]\\ &\quad\times D^{(j,0)}([p])_{m,-j}D^{(j,0)}([p'])_{-(m+1),-j}\\ &\quad=2j!(a(p)c(p')-a(p')c(p))^{2j-1}a(p')a(p).\end{aligned}$$

In a similar way we find that $S_{A_1\underline{C}}S_{A_2}{}^{\underline{C}}$ contains the factor $(a(p)c(p')-a(p')c(p))^{2j-1}$.

The results obtained so far for the vector and tensor current can be generalized easily to an arbitrary current

$$\Phi^{(k,l)}_{A_1\cdots A_{2k}\dot B_1\cdots\dot B_{2l}},$$

$$k+l=\text{integer number},\quad k,l=0,\tfrac{1}{2},1,\cdots.$$

For such a tensor current, expressions similar to those in (15), (16), (22), (23), or (24) will arise where the summation over $\underline{C}$ will run over the

$$2j-(k+l)$$

indices and more. Then the sum

$$S_{A_1\cdots A_n\underline{C}}S_{A_{n+1}A_{2n}}{}^{\underline{C}},$$
$$\underline{C}=C_1\cdots C_{2j-n},\quad 0\leqslant n\leqslant k+l, \tag{29}$$

contains the factor

$$(a(p)c(p')-a(p')c(p))^{2j-k-l}. \tag{30}$$

If we integrate (29) over $\mathbf{x}$ to estimate the charge contribution we shall find that this contribution vanishes as soon as

$$j>\tfrac{1}{2}(k+l),$$

by virtue of (30).

Moreover, all helicity preserving sesquilinear expressions will vanish for $|h|=j>\frac{1}{2}(k+l)$, as required by the Weinberg–Witten theorem.

It is clear that the sesquilinear forms summed over $2j-(l+k)$ indices $\underline{C}$ are crucial for the evaluation of the current and charge contributions. Thus we succeeded not only in proving the Weinberg–Witten theorem in a different manner than was done previously,[1,2] but we were also able to extend it to currents that flip the helicity by showing that, although such a current does not need, in general, to vanish, its contribution to the charge vanishes as soon as (1) is satisfied.

Notice that to get the results presented above it was not necessary to invoke the local conservation law of the currents under consideration; it is only involved implicitly as a necessary condition for a current to induce a charge.

IV. SPINORIAL CHARGES

The method presented in Secs. I–III can be extended to cover also the case of spinorial charges ($k+l=$ half-integer number) under the proviso that the original sesquilinear forms used, e.g., in (11), (12), (17), or (18), have to be replaced by quadratic expressions in at least two fields of which one is bosonic and the other fermionic; one of the relevant terms would be, e.g.,

$$\lim_{y\to x}:T_{A_1\cdots A_{2k}}{}^{C_1\cdots C_{2j}}{}_{\dot{B}_1\cdots\dot{B}_{2l}}{}^{D_1\cdots D_{2j\pm1}}S_{C_1\cdots C_{2j}}S'_{D_1\cdots D_{2j\pm1}}.$$

For this particular case one of the expressions of crucial importance for our calculation would be

$$S_{A_1\cdots A_{(k+l\mp1/2)}C_{(k+l+1\mp1/2)}\cdots C_{2j}}S'_{A_{(k+l+1\mp1/2)}\cdots A_{2(k+l)}}$$
$$\times C_{(k+l+1\mp1/2)\cdots C_{2j}}.$$

Charges which change, e.g., scalar fields into vector fields, etc., can also be tackled with this method, but this kind of charge will probably appear only in a free field theory and therefore is of practically no interest.

V. AN EXAMPLE OF THE ENERGY MOMENTUM TENSOR IN THE ELECTROMAGNETIC FIELD THEORY

As an illustration of our previous considerations, which will perhaps help to make them clearer, let us inspect the case of the free electromagnetic field ($j=|h|=1$), viz.,

$$F_{AC\dot{B}\dot{D}}=(\sigma^{\mu})_{A\dot{B}}(\sigma^{\nu})_{C\dot{D}}F_{\mu\nu}=S_{AC}\epsilon_{\dot{B}\dot{D}}+(S_{BD})^{+}\epsilon_{AC},$$

where S_{AC} creates a particle of helicity (-1) and annihilates a particle of helicity 1.

The energy momentum tensor [i.e., $k=l=1$, hence $\frac{1}{2}(k+l)=j$]

$$T_{\mu\nu}=\,:(-F_{\mu\lambda}F^{\lambda}_{\nu}+\tfrac{1}{4}\eta_{\mu\nu}F_{\lambda}F^{\lambda}):$$

written in the van der Waerden notation reads as

$$\begin{aligned}T_{A_1A_2\dot{B}_1\dot{B}_2}=\tfrac{1}{2}:[&-S_{A_1C}S_{A_2}{}^{C}\epsilon_{\dot{B}_1\dot{B}_2}-S^{+}{}_{\dot{B}_1\dot{C}}S^{+}{}_{\dot{B}_2}{}^{\dot{C}}\epsilon_{A_1A_2}\\&+\tfrac{1}{2}\epsilon_{A_1A_2}\epsilon_{\dot{B}_1\dot{B}_2}(S_{C_1C_2}S^{C_1C_2}+S^{+}{}_{\dot{C}_1\dot{C}_2}S^{+\dot{C}_1\dot{C}_2})\\&+2S_{A_1A_2}S^{+}{}_{\dot{B}_1\dot{B}_2}]:.\end{aligned} \tag{31}$$

The terms occurring in (31) are (23b), (24a), the Hermitian conjugate to them, and (19a). The first four terms in (31), which flip the helicity, do not contribute to the energy momentum operator

$$P_{\mu}=\int d^3x\,T_{0\mu}$$

when restricted to one-particle Hilbert space. The only contribution comes from the last term, which does not flip the helicity.

ACKNOWLEDGMENTS

The author thanks Dr. Ivo Białynicki-Birula, Dr. Anton Z. Capri, and Dr. Helmut Reeh for interesting discussions. The author is also grateful to Dr. Anton A. Capri and Dr. H. Umezawa for the warm hospitality extended to him during his stay at the University of Alberta in Edmonton.

[1] S. Weinberg and E. Witten, Phys. Lett. B **96**, 59 (1980).
[2] J. Łopuszański, J. Math. Phys. **25**, 3503 (1984).
[3] K. M. Case and S. G. Gasiorowicz, Phys. Rev. **125**, 1055 (1962).
[4] See, e.g., R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).
[5] E. C. G. Sudarshan, Phys. Rev. D **24**, 1591 (1981).
[6] S. Coleman and J. Mandula, Phys. Rev. **159**, 1251 (1967).
[7] To prove that the product of self- and anti-self-dual antisymmetric tensor fields vanishes, consider

$$\tilde{S}_{\mu\nu}\equiv(i/2)\epsilon_{\mu\nu\varkappa\lambda}S^{\varkappa\lambda}=S_{\mu\nu}$$

and

$$\tilde{A}_{\mu\nu}=-A_{\mu\nu}.$$

Then

$$S_{\mu\nu}A^{\mu\nu}=-S_{\mu\nu}\tilde{A}^{\mu\nu}=-\tilde{S}_{\mu\nu}A^{\mu\nu}=-S_{\mu\nu}A^{\mu\nu}.$$

Existence and uniqueness of the pressure profile $p(A)$ of a current-carrying quasineutral equilibrium plasma

Michael Kiessling
Institut für Theoretische Physik IV, Ruhr-Universität Bochum, D-4630 Bochum 1, Federal Republic of Germany

(Received 17 November 1987; accepted for publication 13 January 1988)

It is proved that the nonlinear integral equation (quasineutrality condition) for the electrostatic potential ϕ of an isothermal quasineutral current-carrying equilibrium plasma with translational invariance has a solution $\phi(A)$ determined uniquely up to an arbitrary additive constant, where A is the magnetic flux function. This implies the existence and uniqueness of the pressure profile $p(A)$. The existence proof is constructive, which implies its potential usefulness for the practical computation of current-carrying plasma equilibria.

I. INTRODUCTION

Let us consider a fully ionized overall neutral [Eq. (31) below] multispecies plasma with translational invariance (z direction). The plasma shall be in static but current-carrying equilibrium, the current flowing only along the invariant direction. Such systems are frequently discussed as zeroth-order approximations to more realistic geometries both in thermonuclear fusion and in space plasma research. For example, toroidal plasmas, encountered typically in tokamaks, are often replaced by cylindrical plasmas in the large aspect ratio limit.[1–3] In space plasma physics, solar preflare structures[4–6] and planetary magnetospheres[7–9] are often studied theoretically in terms of translationally invariant (two-dimensional) models. It is well-known that then the magnetic induction $\mathbf{B}$ can be represented via

$$\mathbf{B} = \nabla A \times \mathbf{e}_z + B_z \mathbf{e}_z, \tag{1}$$

where B_z is constant and A is the flux function. In the Coulomb gauge, $\partial A/\partial z = 0$. The flux function is related to the z component of the current density j through

$$-\Delta A = \mu_0 j. \tag{2}$$

Now, the interesting point is that unlike in standard magnetostatic theory[10] the current density j cannot be prescribed as an arbitrary function of the space variables. Rather, j has to fulfill

$$j = \frac{dp(A)}{dA}, \tag{3}$$

together with (2), as a necessary and sufficient condition for magnetostatic equilibrium of a quasineutral plasma.[11,12] In (3), p is the kinetic pressure which, in the simplest case, is written as a function of A. In more general cases, p is a nonlocal operator acting on A (see below). This means that in general (2) is a nonlinear partial integrodifferential equation. As a typical feature multiple solutions of (2) might exist for the same boundary conditions for A, or even no solution at all. This property finds an interesting application in the theory of eruptive processes in plasmas, in that spontaneous activity of the plasma may be interpreted as a transition from one equilibrium state to another, or as a loss of equilibrium at a point where equilibrium solutions cease to exist.[6,9,13–15]

The equilibrium condition $p = p(A)$ is a rather broad mathematical statement. From a physicist's standpoint it is of interest to know what special form of $p(A)$ applies to a certain situation. An answer to that problem, however, cannot come from magnetostatic theory alone. Indeed, within the rather coarse macroscopic picture of static ideal magnetohydrodynamics (MHD) any form of $p(A)$ is acceptable. Typical choices are[3,9] $p \sim \exp(A/A_0)$ or[7] $p \sim A^2$, which are made for mathematical convenience. In order to understand the physical meaning of a specific $p(A)$ one needs a more refined description than MHD. A better choice is the microscopic picture of Vlasov theory.

When working within collisionless kinetic theory (Vlasov theory), in principle $p(A)$ is obtained as follows (e.g., Refs. 3 and 8, and references therein). First the equilibrium distribution function is chosen. For a current-carrying plasma with translational invariance any function $F(H,P_z)$ of the one-particle Hamiltonian H and the z component of the one-particle canonical momentum P_z is a solution of the stationary collisionless Boltzmann equation, i.e., a one-particle equilibrium distribution function. From such an equilibrium distribution function one obtains the pressure p in a form $p(A,\phi)$, where ϕ is the electrostatic potential. The electrostatic potential ϕ, on the other hand, depends on A via the quasineutrality condition (see below), which generally is a nonlinear integral equation. Clearly, multiple solutions $\phi_i(A)$ might exist, which can lead to different expressions $p_i(A)$ for the same F. In principle one cannot exclude the possibility that in certain cases no solution $\phi(A)$ of the quasineutrality equation might exist. (This would mean that the chosen F would not correspond to a quasineutral plasma). This leads to the following interesting questions: Given an equilibrium distribution function $F(H,P_z)$, does the quasineutrality equation have a solution $\phi(A)$ such that $p(A)$ exists? If so, is the pressure p uniquely determined in the form $p(A)$, or does the same F lead to several distinct forms $p_i(A)$?

In this paper the special but important case of isothermal systems is discussed. First the basic equations for current-carrying isothermal multispecies plasmas[16,17] are presented. It is then shown that the quasineutrality condition for an isothermal plasma consisting of electrons and an arbitrary number of positively charged ion species determines $\phi(A)$ uniquely up to an arbitrary and irrelevant additive constant. As a consequence, $p(A)$ is uniquely defined for

 0022-2488/88/051258-06$02.50

these systems. The proof is constructive and utilizes some ideas about supersolutions and monotone iterations in Banach spaces.[18] Some concluding remarks are given at the end of this paper.

II. THE BASIC EQUATIONS

In isothermal equilibrium a current-carrying plasma is described by means of a generalized canonical distribution function, which can formally be evaluated by introducing the ideal-gas approximation (for details, see Refs. 16 and 17). The local pressure p then reads

$$p=\sum_s \widehat{N}_s k_B T \frac{\exp(-q_s\beta(\phi-u_sA))}{\int_D \exp(-q_s\beta(\phi-u_sA))d^2r}. \tag{4}$$

In (4), $s=1,2,...,\sigma$ denotes the particle species. Here $\widehat{N}_s$ is the number of particles per unit length along the z direction in the domain $D\subset \mathbb{R}^2$, pertaining to species s. The charge of a particle is denoted by q_s, u_s denotes the average velocity in the z direction, and $\beta=(k_B T)^{-1}$ is the inverse temperature, essentially. For the sake of concreteness the domain $D\subset\mathbb{R}^2$ is assumed to be finite.

From (4) one obtains the local charge density ρ and the current density in the z direction j via partial differentiation,

$$-\rho=\frac{\partial p}{\partial \phi}, \tag{5a}$$

$$j=\frac{\partial p}{\partial A}, \tag{5b}$$

such that in quasineutral equilibrium

$$0=\frac{\partial p}{\partial \phi}, \tag{6a}$$

$$-\mu_0^{-1}\Delta A=\frac{\partial p}{\partial A}. \tag{6b}$$

Note that (6a) implies

$$\frac{\partial p}{\partial A}=\frac{dp}{dA}. \tag{7}$$

Explicitly, the quasineutral equilibria are then described by means of the following two equations:

$$0=\sum_s \widehat{N}_s q_s \frac{\exp(-q_s\beta(\phi-u_sA))}{\int_D \exp(-q_s\beta(\phi-u_sA))d^2r}, \tag{8}$$

$$-\mu_0^{-1}\Delta A=\sum_s \widehat{N}_s q_s u_s \frac{\exp(-q_s\beta(\phi-u_sA))}{\int_D \exp(-q_s\beta(\phi-u_sA))d^2r}, \tag{9}$$

supplemented by appropriate boundary conditions for A. A convenient choice are Dirichlet boundary conditions which, however, are physically determined only up to an arbitrary additive constant. Furthermore, if A and ϕ are a system of solutions of (8) and (9) then so are $A+\widehat{A}$ and $\phi+\widehat{\phi}$, where $\widehat{A}$ and $\widehat{\phi}$ are arbitrary constants [gauge invariance of (8) and (9)].

III. EXISTENCE AND UNIQUENESS OF $p(A)$

In this section it is shown that, for any given well-behaved function A, and provided the arbitrary and irrelevant additive constant $\widehat{\phi}$ is fixed by any convenient gauge [e.g., Eq. (15) below], (8) has a unique solution $\phi(A)$. Since (4) is not altered if a constant is added to ϕ (i.e., choosing another gauge), this results in a unique form $p(A)$ for the pressure p. The proof is divided into two parts: fixing $\widehat{\phi}$ by means of a convenient gauge, (1) it is shown that there exists at least one solution $\phi(A)$ of (8); and (2) it is shown that that solution is unique.

The idea is as follows: Since the normalizing integrals in (8) are independent of position in space we can treat them as σ unknown parameters $U_s\in\mathbb{R}^+$ that have to be determined self-consistently *a posteriori*. This means that we can split (8) into a system of equations: (1) a transcendental equation for ϕ, parametrized by the normalizing integrals, and (2) a system of σ equations for the normalizing integrals. First it is shown that there exists a unique solution $\phi(A;\{U\})$ of the transcendental equation for ϕ, where $\{U\}\in\mathbb{R}^{\sigma,+}$ denotes the normalizing integrals, arranged as an element of the positive cone $\mathbb{R}^{\sigma,+}$ of the Banach space $\mathbb{R}^\sigma$. This solution $\phi(A;\{U\})$ has to be inserted into the system of equations for the normalizing integrals, which is treated as a nonlinear fixed-point equation in $\mathbb{R}^\sigma$. It is then shown that this equation has a solution. The uniqueness (except for $\widehat{\phi}$) of the solution of (8) is shown, finally, by proving that (8) defines the stationary points of a convex functional.

In this paper the investigations are restricted to nonsingular flux functions A, such that all normalizing integrals have finite values. Although from elliptic regularity theory[19] we know that the nonsingular physical solutions of (2) are at least $C^1(D)$, for our purposes it suffices to postulate that A is bounded only. The case of singular A, which is of physical interest, too, will be treated elsewhere.

Proof. (a) Existence of a solution: Let A be bounded. Under the assumption that ϕ is bounded, which is verified below, the normalizing integrals exist. Let us introduce the abbreviations

$$\int_D \exp(-q_s\beta(\phi-u_sA))d^2r\equiv U_s. \tag{10}$$

Note that the q_s are integer multiples of the elementary charge e. Furthermore, in fully ionized plasmas the temperatures are generally that high that the negatively charged particles are electrons, exclusively. We therefore consider a plasma consisting of electrons ($q_e=-e$) and $\sigma-1$ positively charged ion species i, with $q_i=z_ie$ and $z_i\in\mathbb{N}$. We then introduce the further abbreviations

$$\xi_\phi\equiv\exp(e\phi/k_BT) \tag{11}$$

and

$$g_i(A)\equiv\frac{\widehat{N}_i}{\widehat{N}_e}\frac{U_e}{U_i}\exp\left(\frac{u_e+z_iu_i}{k_BT}eA\right). \tag{12}$$

Now we write (8) in the form

$$\xi_\phi^{z^*+1}-\sum_{z=1}^{z^*} z\xi_\phi^{z^*-z}\sum_{\substack{i\\ z_i=z}} g_i(A)=0, \tag{13}$$

where

$$z^*\equiv\max\{z_i\}. \tag{14}$$

The left-hand side of (13) is a polynomial in ξ_ϕ of degree z^*+1, which we call $P(\xi_\phi)$. We need to look for a positive solution ξ_ϕ of (13) only.

There exists a unique positive solution ξ_ϕ of (13). This follows from the fact that the coefficient in front of the highest power is 1, i.e., positive, whereas all other coefficients are negative [since $g_i(A) > 0$], and from the following Lemma.

Lemma: "Let ξ be a real variable and $P(\xi) = \Sigma_{i=0}^{k} p_i \xi^i$ a real valued polynomial of degree $k \geqslant 1$, with the property

(P): $p_k > 0, p_0 < 0$, and $p_i \leqslant 0$, for $0 < i < k$.

Then $P(\xi)$ has a unique positive zero."

The proof of that Lemma is given in Appendix A. Since $P(\xi_\phi)$ has the property (P) of the above Lemma, and since furthermore (11) is uniquely invertible it follows immediately that there exists a unique solution $\phi(A;\{U\})$ of (13), provided $\{U\}$ is finite. Obviously $\phi(A;\{U\})$ is bounded. In many relevant cases the positive zero of (13) can be calculated analytically.

It remains to show that upon inserting this parametrized solution $\phi(A;\{U\})$ into the left-hand side of (10) there exists at least one solution $\{U\}$, with $U_i \neq 0$ for all i, of the resulting system of equations. Since the normalizing integrals are non-negative, the F_i are bounded below by 0. As a consequence the system of equations for $\{U\}$ constitutes a nonlinear fixed-point problem in the positive cone $\mathbb{R}^{\sigma,+}$ of the Banach space $\mathbb{R}^\sigma$ (e.g., Ref. 18).

We make use of the gauge freedom for ϕ and fix $\hat{\phi}$ by choosing

$$U_e \equiv 1, \tag{15}$$

which is always possible if A is well-behaved (lengths are measured in dimensionless units). We are then left with the nonlinear system of $\sigma - 1$ equations

$$U_i = F_i(\{U\}) \tag{16}$$

supplemented by (15). The F_i are defined by the left-hand side of (10), where we have inserted $\phi(A;\{U\})$ for ϕ. We have thus reduced our fixed-point problem in $\mathbb{R}^{\sigma,+}$ to a fixed-point problem in the affine space $1 \oplus \mathbb{R}^{\sigma-1}$. We now show that there exists a monotonously decreasing iterational sequence $\{\{U\}^{(n)}\} \in 1 \oplus \mathbb{R}^{\sigma-1,+}$, where $\mathbb{R}^{\sigma-1,+}$ is the positive cone of $\mathbb{R}^{\sigma-1}$. Since the F_i are bounded below, that sequence must converge. The limit is in $1 \oplus \mathbb{R}^{\sigma-1,+}$ (Appendix B). This then proves the existence of a bounded $\phi(A)$.

It is readily shown that

$$\frac{\partial P}{\partial U_i} > 0, \tag{17}$$

for all i. This together with property (P) of $P(\xi_\phi)$ implies that the zero of $P(\xi_\phi)$ is shifted toward smaller values with increasing U_i. This again by means of (11) implies

$$\frac{\partial \phi(A;\{U\})}{\partial U_i} < 0, \tag{18}$$

for all i, and consequently

$$\frac{\partial F_i}{\partial U_j} > 0, \tag{19}$$

for all i and j, which follows from the definition of the F_i. Property (19) and the boundedness from below of the F_i imply[18] that, in the case where there exist $U_i^{(0)}$ such that

$$U_i^{(0)} > F_i(\{U\}^{(0)}), \tag{20}$$

for all i simultaneously, the sequences

$$U_i^{(n+1)} = F_i(\{U\}^{(n)}) \tag{21}$$

converge monotonously, when starting with a $\{U\}^{(0)}$ fulfilling (20). [A $\{U\}^{(0)}$ with the property (20) is called a strict supersolution.] The existence of such $U_i^{(0)}$ can be shown as follows. We observe that $\tilde{\phi}(A;\{U\})$, solving

$$\xi_\phi^{z^*+1} + P(0) = 0 \tag{22}$$

with

$$P(0) = -z^* \sum_{\substack{j \\ z_j = z^*}} g_j(A) < 0 \tag{23}$$

is pointwise smaller than $\phi(A;\{U\})$. (Note further that $\tilde{\phi}$ is independent of all U_j for which $z_j \neq z^*$.) Inserting $\tilde{\phi}(A;\{U\})$ instead of $\phi(A;\{U\})$ into the left-hand side of (10) yields upper bounds $\widetilde{F}_i$ for the F_i. It is then easy to construct $U_i^{(0)}$ with the property

$$U_i^{(0)} > \widetilde{F}_i(\{U\}^{(0)}), \tag{24}$$

which then also fulfill (20). Indeed, it suffices to consider the special case where all $U_i^{(0)}$ have the same value, W. We then have to find a W that obeys

$$W > \widetilde{F}_i(\{W\}), \tag{25}$$

for all i. When written explicitly, (25) reads

$$W > W^{z_i/(z^*+1)} \Gamma_i, \tag{26}$$

for all i. The Γ_i are fixed numerical factors (independent of W) and read

$$\Gamma_i = \int_D \exp(z_i e\beta u_i A) \Bigg[z^* \sum_{\substack{j \\ z_j = z^*}} \hat{N}_j \hat{N}_e^{-1} \times \exp\left(e\beta(u_e + z_j u_j)A\right) \Bigg]^{-z_i/(z^*+1)} d^2 r. \tag{27}$$

Because of (14) the right-hand side of (26) increases slower than linearly with W. This means that a supersolution is given by choosing

$$W > \max_i \{\Gamma_i^{(z^*+1)/(z^*+1-z_i)}\}. \tag{28}$$

This completes the existence part of the proof.

Proof. (b) Uniqueness of the solution: It suffices to show that the solutions of (8) define the stationary points of a strictly convex functional (over a suitable function space). The quasineutrality condition (8) can be derived from a variational principle

$$\delta_\phi J = 0 \tag{29}$$

with

$$J[\phi] = \sum_s \hat{N}_s k_B T \ln \int_D \exp(-q_s \beta(\phi - u_s A)) d^2 r. \tag{30}$$

This means that J is stationary for ϕ being a solution of (8). Because of the overall neutrality of the plasma, i.e.,

$$\sum_s \hat{N}_s q_s = 0, \tag{31}$$

which follows from (8), the functional $J[\phi]$ has the invariance property

$$J[\phi + \hat{\phi}] = J[\phi] \tag{32}$$

for $\hat{\phi} = \text{const}$, which is easily shown. Moreover, J is convex: Consider the straight line in function space

$$\phi_\lambda \equiv (1-\lambda)\phi^{(1)} + \lambda\phi^{(2)}, \tag{33}$$

where $\phi^{(1)}$ and $\phi^{(2)}$ are bounded and $\lambda \in \mathbb{R}$. Inserting (33) for ϕ into (30) yields a function of λ,

$$J[\phi_\lambda] = G(\lambda). \tag{34}$$

Since ϕ_λ and A are bounded, $G(\lambda)$ is an arbitrarily differentiable function. The second derivative of G reads

$$\frac{d^2G}{d\lambda^2} = \sum_s \hat{N}_s q_s^2 \beta \langle (\psi - \langle\psi\rangle_s)^2\rangle_s, \tag{35}$$

where

$$\psi \equiv \phi^{(2)} - \phi^{(1)}. \tag{36}$$

The angular brackets denote the averages

$$\langle f\rangle_s = \frac{\int_D f \exp(-q_s\beta(\phi - u_s A))d^2r}{\int_D \exp(-q_s\beta(\phi - u_s A))d^2r} \tag{37}$$

with $\phi = \phi_\lambda$. Obviously

$$\frac{d^2G}{d\lambda^2} \geqslant 0. \tag{38}$$

The equality sign in (38) holds iff $\psi \equiv \hat{\phi} = \text{const}$, consistent with (32). Result (38) means that J is convex over the function space of real-valued bounded functions on D, and strictly convex over any subspace for which $\psi \neq \text{const}$. Fixing $\hat{\phi}$ by means of a convenient gauge, e.g., (15), this implies[19,20] that there can exist at most one solution of (8).

That completes the proof.

IV. CONCLUDING REMARKS

It has been shown that in the special but important case of isothermal plasmas consisting of electrons and several positively charged ion species there exists a unique form $p(A)$ for the pressure p. The restriction to one negatively charged species was made for convenience, and that seems to be the most relevant case. The inclusion of more than one negatively charged particle species is no problem and requires only moderate changes. (Note that the uniqueness proof makes no use of the assumption that only electrons are present.) Future investigations have to show what conditions more general distribution functions must fulfill in order to share that property with the isothermal systems. In that sense this paper is a first step in that direction.

Concerning the directly related problem of finding solutions of the system (8) and (9) in practice, the proof given in Sec. III is of value, too, since the proof is constructive. For example, consider a hydrogen plasma, i.e., a plasma consisting of electrons (index e) and protons (index p) only. The solution of (8), or (13), then reads

$$\phi(A) = [(u_p + u_e)/2]\, A + \hat{\phi}. \tag{39}$$

This leads to

$$p(A) = 2\hat{N}_e k_B T \frac{\exp(A/A_0)}{\int_D \exp(A/A_0) d^2r}, \tag{40}$$

where

$$A_0 = 2(\beta e(u_p - u_e))^{-1}. \tag{41}$$

This results[17] in

$$-\Delta A = \mu_0 e(u_p - u_e)\hat{N}_e \frac{\exp(A/A_0)}{\int_D \exp(A/A_0)d^2r}. \tag{42}$$

Thus the relation $p(A) \sim \exp(A/A_0)$, which is a common choice in MHD (see Sec. I), comes out as a special case of Eqs. (8) and (9). For alternative approaches concerned with a derivation of that relation between p and A, see Refs. 21 and 22.

As a second, more complicated example, consider a three species plasma consisting of electrons (index e), protons (index p), and a further singly charged ion species (index i). Because of the gauge invariance of (8) and (9) (cf. the remark at the end of Sec. II) we can choose a special gauge such that the three normalizing integrals $U_{e,p,i}$ have the same value U. The solution of (8) is then

$$\phi(A) = (k_B T/2e)\ln(\nu \exp(e\beta(u_p + u_e)A) + (1-\nu)\exp(e\beta(u_i + u_e)A)), \tag{43}$$

where $\nu = \hat{N}_p/\hat{N}_e$ and $1 - \nu = \hat{N}_i/\hat{N}_e$. This results in[23]

$$p(A) = 2\hat{N}_e k_B T U^{-1}(\nu \exp(2A/A_0) + (1-\nu)\exp(2\alpha A/A_0))^{1/2}, \tag{44}$$

where

$$\alpha = (u_i - u_e)/(u_p - u_e). \tag{45}$$

In the limit $\nu \to 1$ this expression reduces to (40), as should be the case.

It can be expected that because of the complexity of (8) and (9) explicit analytical solutions will be possible only in exceptional cases.[17,23] In general a numerical treatment will be required. In cases where (8) can be solved in closed form for $\phi(A)$, e.g., Eqs. (39) and (43), the numerical effort is at least reduced as compared to the cases where $\phi(A)$ cannot be obtained explicitly analytically. In the latter cases, the uniqueness-existence proof for $\phi(A)$ given here provides an algorithm that might play a useful role as part of a general numerical solution procedure of (8) and (9).

ACKNOWLEDGMENTS

I like to thank K. Schindler and H. J. Ziegler for a critical reading of the manuscript.

This work was supported by the Deutsche Forschungsgemeinschaft through the Sonderforschungsbereich Plasmaphysik Bochum/Jülich.

APPENDIX A: PROOF OF THE LEMMA

Proof: The existence of a positive zero follows immediately from

$$\lim_{\xi = \infty} P(\xi) = +\infty \tag{A1}$$

and

$$P(0) = p_0 < 0 \tag{A2}$$

and the continuity of polynomials. The uniqueness of that zero can be shown, for instance, by *reductio ad absurdum*: Assume there exist at least two positive zeros of $P(\xi)$. Then the mean-value theorem of elementary calculus in connection with property (P) guarantees that the derivative $P'(\xi)$ has at least three positive zeros. Again due to (P) the second derivative $P''(\xi)$ must have three positive zeros and so on until we arrive at the statement that a polynomial of degree two must have three positive zeros, which is impossible. That proves the Lemma.

APPENDIX B: PROOF OF $U_i^{(\infty)}>0$

A positive limit of (21) for all i is a necessary and sufficient criterion for a bounded $\phi(A)$ to exist. The boundedness from below of the F_i by 0 suffices to prove convergence of (21), starting with $U_i^{(0)} = W$ [Eq. (28)] for all i, but is insufficient to prove a positive limit of (21). If there would exist positive lower bounds of the F_i it would immediately follow that a positive limit of (21) exists; however, 0 is the infimum of the F_i. Hence we need another idea.

The existence of a positive limit of (21) can be established[18] by showing that there exist subsolutions in addition to the supersolutions. In the following, we denote a subsolution of (16) by $\underline{U}_i$ and a supersolution by $\overline{U}_i$. A $\underline{U}_i$ is called a subsolution iff

$$\underline{U}_i \leqslant F_i(\{\underline{U}\}), \tag{B1}$$

whereas a supersolution $\overline{U}_i$ fulfills

$$\overline{U}_i \geqslant F_i(\{\overline{U}\}) \tag{B2}$$

[cf. Eq. (20) and text below (21)]. If subsolutions and supersolutions of (21) exist simultaneously for all i, with $\underline{U}_i \leqslant \overline{U}_i$, the following relations hold[18]:

$$\underline{U}_i \leqslant U_i^{(\infty)} \leqslant \overline{U}_i, \tag{B3}$$

for all i. The existence of positive $\underline{U}_i$ is shown below. [Note that 0 is a subsolution of (16); however, it corresponds to an everywhere unbounded "solution" $\phi(A)$, which is unphysical.]

Proof: It suffices to prove the existence of any special subsolution, i.e., a subsolution having any convenient form we like. Therefore, in the following we introduce a parameter $V \in \mathbb{R}^+$ and consider a one-parameter sequence $\{U\}(V) \in 1 \oplus \mathbb{R}^{\sigma-1,+}$ with the special form

$$U_i(V) = \Theta_z(V), \quad \text{for } z_i = z, \tag{B4}$$

with $\Theta_z(V) \in \mathbb{R}^+$. We assume further that

$$\Theta_z(V) \to 0^+, \quad \text{for } V \to 0^+ \tag{B5}$$

for all z. The $\Theta_z(V)$ will be specified below. We now show that one can find subsolutions $\{\underline{U}\}(V)$ for $V \to 0^+$, provided the $\Theta_z(V)$ are chosen suitably.

The F_i are now obtained by means of the solution $\xi_\phi^{(V)}$ of

$$P^{(V)}(\xi_\phi) = 0, \tag{B6}$$

where $P^{(V)}(\xi_\phi)$ is the polynomial $P(\xi_\phi)$ as given by the left-hand side of (13) for $\{U\} = \{U\}(V)$, as given by (B4). We have

$$P^{(V)}(\xi_\phi) = \xi_\phi^{z^*+1} - \sum_{z=1}^{z^*} z\xi_\phi^{z^*-z}(\Theta_z(V))^{-1} \sum_{\substack{i \\ z_i = z}} h_i(A), \tag{B7}$$

where the h_i are defined by

$$h_i(A) = U_i g_i(A). \tag{B8}$$

Because of (12), the h_i are independent of the U_i, i.e., independent of V. Equation (B5) guarantees that the solution $\xi_\phi^{(V)}$ tends to infinity as $V \to 0^+$. We now fix the $\Theta_z(V)$ by setting

$$\xi_\phi^{(V)}(A) = V^{-a}\eta(A) \tag{B9}$$

with $a>0$ (see below), and obtain

$$\Theta_z(V) = \theta_z V^{a(z+1)}, \tag{B10}$$

where the θ_z are arbitrary positive constants (independent of V). Obviously, (B5) is fulfilled. For simplicity we choose

$$\theta_z = 1, \tag{B11}$$

for all z. The function $\eta(A)$ is the unique solution (cf. Appendix A) of

$$\eta^{z^*+1} - \sum_{z=1}^{z^*} z\eta^{z^*-z} \sum_{\substack{i \\ z_i = z}} h_i(A) = 0. \tag{B12}$$

The $F_i(\{U\}(V))$ are then given by

$$F_i(\{U\}(V)) = \kappa_i V^{az_i}, \tag{B13}$$

where the κ_i are fixed numerical factors, given by

$$\kappa_i = \int_D \exp(z_i e\beta u_i A)(\eta(A))^{-z_i}\, d^2r. \tag{B14}$$

In order to obtain subsolutions we now have to find V such that

$$\Theta_z(V) \leqslant F_i(\{U\}(V)) \tag{B15}$$

for all i with $z_i = z$ and for all z simultaneously. By means of (B10), (B11), and (B13) the V have to fulfill

$$V^{a(z+1)} \leqslant \kappa_i V^{az}, \tag{B16}$$

for all i with $z_i = z$ and for all z simultaneously. Clearly, (B16) is fulfilled for

$$V \leqslant (\min_i \kappa_i)^{1/a}, \tag{B17}$$

where a is indeed arbitrary, provided $a>0$. For simplicity, one can choose $a=1$.

That proves the existence of a positive limit of (21) for each i.

A final remark seems necessary, concerning the overall neutrality of the plasma [Eq. (31)]. Note that (31) did not enter the existence proof of a fixed point in $1 \oplus \mathbb{R}^{\sigma-1,+}$ of (16). On the other hand, a plasma with a net total charge, i.e., for which (31) does not hold, cannot be quasineutral since an everywhere vanishing charge density cannot generate a net charge, which is trivial. It is indeed easily verified that, upon integrating (8), Eq. (31) is a necessary condition for (8) to have a solution. This seemingly paradoxical situation is resolved by noting Eq. (15). Let the U_i be solutions of (16), obtained with U_e replaced by 1 everywhere in the g_i [Eq. (12)]. Upon inserting the corresponding $\phi(A;\{U\})$ into the left-hand side of (10), Eq. (15) then yields an equation for the $\widehat{N}_s$. It can be readily shown that (15) can be fulfilled if and only if (31) holds.

[1]W. H. Bennett, Phys. Rev. **45**, 890 (1934).
[2]V. D. Shrafranov, J. Nucl. Energy **II**, 86 (1957).
[3]O. Ågren and H. Persson, Plasma Phys. Control. Nucl. Fusion **26**, 1177 (1984).
[4]B. C. Low, Astrophys. J. **197**, 251 (1975).

[5]B. C. Low, Astrophys. J. **212**, 243 (1977).
[6]J. Birn, H. Goldstein, and K. Schindler, Sol. Phys. **57**, 81 (1978).
[7]G.-H. Voigt and R. A. Wolf, J. Geophys. Res. **80**, 4046 (1985).
[8]K. Schindler and J. Birn, Phys. Rep. **47**, 109 (1978).
[9]W. Zwingmann, J. Geophys. Res. **88**, 9101 (1983).
[10]J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), 2nd ed.
[11]H. Grad and H. Rubin, in *Proceedings of the Second Internatinal Conference on the Peaceful Uses of Atomic Energy* (United Nations, Geneva, 1958), Vol. 31, p. 190.
[12]D. Shafranov, in *Reviews of Plasma Physics*, edited by M. A. Leontovich (Consultants Bureau, New York, 1966), Vol. 2, p. 103.
[13]J. Heyvaerts, J. M. Lasry, M. Schatzmann, and G. Witomsky, Lecture Notes Math. **782**, 160 (1980).
[14]E. R. Priest, *Solar Magnetohydrodynamics* (Reidel, Dordrecht, 1982).
[15]M. Hesse and M. Kiessling, Phys. Fluids **30**, 2720 (1987).
[16]M. Kiessling, R. P. Brinkmann, and K. Schindler, Phys. Rev. Lett. **56**, 143 (1986).
[17]M. Kiessling and K. Schindler, J. Geophys. Res. **92**, 5795 (1987).
[18]H. Amann, J. Diff. Eq. **21**, 363 (1976).
[19]Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds, and Physics* (North-Holland, New York, 1982), 2nd ed.
[20]M. Reed and B. Simon, *Functional Analysis* (Academic, New York, 1980), 2nd ed.
[21]A. Bhattacharjee, J. Geophys. Res. **92**, 4735 (1987).
[22]D. Montgomery, L. Turner, and G. Vahala, J. Plasma Phys. **21**, 239 (1979).
[23]M. Kiessling and H. J. Ziegler, "A theory of self-consistent two-dimensional tail equilibria for quasi-neutral $e^-H^+O^+$ mixtures," submitted to J. Geophys. Res.

Addendum: On Geroch's limit of space-times and its relation to a new topology in the space of Lie groups [J. Math. Phys. 28, 1928 (1987)]

Hans-Jürgen Schmidt
Central Institute for Astrophysics, Academy of Sciences of the GDR, DDR-1591 Potsdam, R.-Luxemburg-Str. 17a, German Democratic Republic

(Received 1 December 1987; accepted for publication 16 December 1987)

My attention has been directed to Refs. 1–3, which should now be honored.

In Ref. 1 §6 ("Examples of kinematic systems") the topology in the space of Lie groups is already defined. Segal gets as a result that a compact semisimple Lie group is not a limiting case of any nonisomorphic Lie algebra.

In Refs. 2 and 3 and many subsequent papers, not the full topology but its restriction to two-point subsets of the space of Lie groups has been considered frequently: A Lie group G can be contracted (or deformed) to another Lie group H, if the constant sequence G converges to H in the Segal topology.[1]

The most often used application of this concept is from Segal[1]: One physical theory is a limiting case of another one if the underlying Lie groups are[2] (e.g., Lorentz group $\rightarrow$ Galilean group); each Lie group can be contracted to the Abelian Lie group, thus each quantum theory possesses a classical limit.[1]

Further, there exist nontrivial relations between the Segal topology and the set of representations of the respective groups.[3]

In the context of the present topic one should also note Refs. 4 and 5 for similar questions within more general algebras and Refs. 6 and 7, where the limits de Sitter $\rightarrow$ Minkowski space-time and mass $\rightarrow 0$ have been considered.

ACKNOWLEDGMENTS

I thank Professor M. A. H. MacCallum for a constructive conversation and for directing my attention to Refs. 2 and 3, and Dr. R. Schimming for making some useful comments.

[1]See I. E. Segal, Duke Math. J. **18**, 221 (1951), for a class of operator algebras determined by groups.
[2]See E. Inonu and E. P. Wigner, Proc. Natl. Acad. Sci. USA **39**, 510 (1953), for the contraction of groups and their representations.
[3]See E. I. Saletan, J. Math. Phys. **2**, 1 (1961), for the contraction of Lie groups.
[4]See M. Gerstenhaber, Ann. Math. **79**, 59 (1964), for the deformation of rings and algebras.
[5]See J. Niederle, Arch. Math. **19**, 145 (1983), for conditions for trivial principal tolerances.
[6]See B. Binegar, C. Fronsdal, and W. Heidenreich, Ann. Phys. (NY) **149**, 254 (1983), for de Sitter QED.
[7]See M. Flato, D. Sternheimer, and C. Fronsdal, Commun. Math. Phys. **90**, 563 (1983), for difficulties with massless particles.

 0022-2488/88/051264-01$02.50

Erratum: The Lorentz group and the Thomas precession. II. Exact results for the product of two boosts [J. Math. Phys. 27, 157 (1986)]

N. A. Salingaros
Division of Mathematics, The University of Texas, San Antonio, Texas 78285

(Received 2 February 1988; accepted for publication 17 Feburary 1988)

In a recent article,[1] Baylis and Jones uncovered two errors in the work[2] of the author. The first error was corrected in an Erratum.[3] The second error, however, has implications of a fundamental nature that need to be cleared up.

Two successive Lorentz boosts by boost vectors **a** and **b** are equivalent to a single boost by a net vector **s**, along with a spatial rotation by the Wigner angle $\boldsymbol{\theta}$ [from Eqs. (12) and (13) of Ref. 2]. The order of operations is from right to left:

$$\mathbb{L}(\mathbf{b}) \vee \mathbb{L}(\mathbf{a}) = \mathbb{R}(\boldsymbol{\theta}) \vee \mathbb{L}(\mathbf{s}), \tag{1a}$$

$$\mathbb{L}(\mathbf{a}) \vee \mathbb{L}(\mathbf{b}) = \mathbb{L}(\mathbf{s}) \vee \mathbb{R}(-\boldsymbol{\theta}). \tag{1b}$$

An additional correction angle appeared to result from calculating the combination (1) in the four-dimensional Clifford algebra[2] compared to the equivalent calculation in a three-dimensional Clifford algebra.[1,4–10] It turns out that there is no additional correction angle, and so $\boldsymbol{\phi} = \mathbf{0}$, $\mathbf{d} = \mathbf{s}$, and $\mathbf{V}_d = \mathbf{V}_s$ in Refs. 2 and 3. It is important to point out that this error was not due to the failure of the Clifford algebra method, but to an oversight by this author.

The situation is now clear, following the checking of the calculations of Ref. 2 by Baylis and Jones.[1] The very disturbing consequences of an additional correction angle mentioned in Ref. 3 are no longer an issue. Combinations of Lorentz transformations do not depend upon the dimensionality of the Clifford algebra. Also, there is exact symmetry in the order of two noncollinear boosts in the sense outlined in Ref. 1.

I would now like to discuss something that did not appear in Ref. 2. A very elegant result, that three boosts can, in a special case, result in no rotation correction, is noted in Ref. 1. The following expression follows immediately from multiplying (1a) and (1b):

$$\mathbb{L}(\mathbf{a}) \vee \mathbb{L}(2\mathbf{b}) \vee \mathbb{L}(\mathbf{a}) = \mathbb{L}(2\mathbf{s}). \tag{2}$$

This result is not generally known, despite its simplicity. The Wigner angle cancels when boosting by the same vector before and after. Nevertheless, the fact that the net boost 2**s** in (2) is given by the standard combination of two boosts **a** and **b** (1) was overlooked; Baylis and Jones[1] calculate instead the analogous expression

$$\mathbb{L}(\tfrac{1}{2}\mathbf{b}) \vee \mathbb{L}(\mathbf{a}) \vee \mathbb{L}(\tfrac{1}{2}\mathbf{b}) = \mathbb{L}(\mathbf{c}), \tag{3}$$

where **c** is not directly related to **s**. The rather complicated expression for **c** [see Eqs. (36) and (37) in Ref. 1] obscures the simple elegance of the result given by Eq. (2).

In concluding, the boost and Wigner rotation parameters are presented again for clarification. Unfortunately, the unit vectors in $\mathbf{V}_s$ were not printed in Eq. (33) of Ref. 2. Also, $\mathbf{V}_s$ is not given explicitly in Ref. 1, and a reader might easily mistake Eq. (27b) of Ref. 1 for $\mathbf{V}_s$. From Eqs. (41), (42b), and (42c) of Ref. 2,

$$\begin{aligned}
\boldsymbol{\theta} &= \frac{\mathbf{V}_a \times \mathbf{V}_b}{|\mathbf{V}_a \times \mathbf{V}_b|}\, 2 \arctan \Phi, \\
\Phi &= \frac{\gamma_a \gamma_b |\mathbf{V}_a \times \mathbf{V}_b|}{1 + \gamma_a + \gamma_b + \gamma_s}, \quad \gamma_s = \gamma_a \gamma_b (1 + \mathbf{V}_a \cdot \mathbf{V}_b), \\
\mathbf{V}_s &= \frac{\mathbf{V}_a + (\mathbf{V}_b/\gamma_a) + (1 - 1/\gamma_a)(\mathbf{V}_b \cdot \hat{\mathbf{V}}_a)\hat{\mathbf{V}}_a}{1 + \mathbf{V}_a \cdot \mathbf{V}_b} \\
&= \frac{\mathbf{V}_a + \mathbf{V}_b + (1 - 1/\gamma_a)(\mathbf{V}_b \times \hat{\mathbf{V}}_a) \times \hat{\mathbf{V}}_a}{1 + \mathbf{V}_a \cdot \mathbf{V}_b}.
\end{aligned} \tag{4}$$

[1] W. E. Baylis and G. Jones, J. Math. Phys. **29**, 57 (1988).
[2] N. A. Salingaros, J. Math. Phys. **27**, 157 (1986).
[3] N. A. Salingaros, J. Math. Phys. **28**, 492 (1987).
[4] A. J. Macfarlane, J. Math. Phys. **3**, 1116 (1962).
[5] H. A. Farach, Y. Aharonov, C. P. Poole, and S. I. Zanette, Am. J. Phys. **47**, 247 (1979).
[6] J. D. Hamilton, Can. J. Phys. **59**, 213 (1981).
[7] C. B. van Wyk, Am. J. Phys. **52**, 853 (1984).
[8] A. Ben-Menahem, Am. J. Phys. **53**, 62 (1985).
[9] M. Rivas, M. A. Valle, and J. M. Aguirregabiria, Eur. J. Phys. **7**, 1 (1986).
[10] A. C. Hirshfeld and F. Metzger, Am. J. Phys. **54**, 550 (1986).

 0022-2488/88/051265-01$02.50