Nonlinear equations with superposition formulas and the exceptional group G_2. I. Complex and real forms of g_2 and their maximal subalgebras

J. Beckers and V. Hussin[a)]
Physique théorique et mathématique, Université de Liège, Institut de Physique au Sart Tilman, B.5. B-4000, Liège 1, Belgium

P. Winternitz
Centre de recherches mathématiques, Université de Montréal, C.P. 6128, Succursale A, Montréal, Québec H3C 3J7, Canada

(Received 12 March 1986; accepted for publication 7 May 1986)

In order to study nonlinear ordinary differential equations with superposition principles, related to the exceptional simple Lie group G_2, the complex and real forms of its Lie algebra are examined and their maximal subalgebras are summarized. In particular the parabolic subalgebras of the noncompact real form $g_2^{\rm NC}(\mathbb{R})$ are determined. Explicit matrix realizations of the fundamental representation $D(1,0)$ are used and studied in connection with invariant subspaces in a seven-dimensional (complex or real) vector space. The results are collected in three tables of specific interest for the study of nonlinear differential equations, which will be developed in Paper II of this series.

I. INTRODUCTION

A series of recent publications has been devoted to the problem of identifying and classifying all systems of first-order nonlinear ordinary differential equations with superposition formulas.[1–6] The equations under study have the form

$$\dot{y}^{\mu} = \sum_{k=1}^{r} Z_k(t)\eta_k{}^{\mu}(\mathbf{y}), \quad \mu = 1,...,n, \tag{1.1}$$

and in this context a superposition formula is a mapping F: $\mathbb{C}^{n(m+1)} \to \mathbb{C}^n$ (or $\mathbb{R}^{n(m+1)} \to \mathbb{R}^n$) expressing the general solution $\mathbf{y}(t)$ of (1.1) in terms of m particular solutions $\mathbf{y}_i(t)$ and n significant constants C_i:

$$\mathbf{y}(t) = \mathbf{F}(\mathbf{y}_1(t),...,\mathbf{y}_m(t),C_1,...,C_n). \tag{1.2}$$

The problem of characterizing nonlinear ordinary differential equations with superposition formulas goes back to Lie.[7] We shall not review the known results here, nor the motivation for our interest in these equations and the explicit superposition formulas.[1–6]

Let us just mention that a system of ODE's with a superposition formula can be associated with every Lie group–Lie subgroup pair $G \supset G_0$. To obtain the equations, one must realize the homogeneous space G/G_0 explicitly and introduce convenient coordinates on this space. It has been shown that cases of particular interest are obtained when G is a simple Lie group and G_0 one of its maximal subgroups.[4]

Attention has so far been focused on the case when G is a classical complex or real Lie group. Use was made of the defining matrix representations of the corresponding simple classical Lie algebras and Lie groups. The homogeneous spaces G/G_0 were constructed as Grassmannians, or in some other form.[1–6]

The purpose of this article is to start the analysis of the Cartan exceptional Lie groups[8,9] G_2, F_4, E_6, E_7, and E_8 in the context of nonlinear superposition formulas. More specifically, we concentrate on the simplest case, namely the complex and real forms of the exceptional Lie group G_2.

It should be emphasized that the group G_2 is of considerable interest in physics, quite independently of the application mentioned above. The real compact form $G_2^{\rm C}(\mathbb{R})$ made its appearance in elementary particle physics in the early 1960's,[10] as a possible candidate for the symmetry group of strong interaction physics and more particularly for hadron spectroscopy [it contains the all-important SU(3) group as one of its maximal subgroups].

Quantum theories based on octonionic quark fields were proposed in the 1970's[11,12] and the exceptional group $G_2^{\rm C}(\mathbb{R})$ was once again examined: this time as the automorphism group[13] of the octonions. Other applications include non-Abelian gauge theories where G_2 occurs as one of the groups that can accommodate three-quark color singlets[14] and to many-body problems in nuclear and atomic physics[15–17] [here $G_2(\mathbb{R})$ figures, e.g., in the group reduction chain $\mathrm{U}(7) \supset \mathrm{O}(7) \supset G_2 \supset \mathrm{SU}(3)$]. More recently, possible global symmetries of extended supergravities[18] have also lead to the use of exceptional Lie groups, in particular G_2 and its fundamental representation. An interesting application of G_2 occurs in the study of a Toda lattice with unequal masses.[19]

The present article is devoted to group theoretical preliminaries, that should be of use in any physical application of the group G_2. It contains some known results and some, to our knowledge, new ones, on the subgroup structure and realizations of the complex group $G_2(\mathbb{C})$, the real compact form $G_2^{\rm C}(\mathbb{R})$, and the real noncompact form $G_2^{\rm NC}(\mathbb{R})$. The sequel (Paper II) will make use of these results to obtain explicitly the nonlinear ODE's with superposition formulas, associated to the various forms of G_2 (see also Ref. 20).

Section II contains some general results on the group G_2 and its complex Lie algebra $g_2(\mathbb{C})$. The two real forms of g_2 and some of their properties are discussed in Sec. III. The parabolic subalgebras[21] of the noncompact real form $g_2^{\rm NC}(\mathbb{R})$ are obtained using a method employed by Cornwell,[22] dealing with the Iwasawa[23] and Langlands[24] decompositions. Finally, Sec. IV is devoted to explicit realizations

a) Chargé de recherches, F.N.R.S., Belgium.

 0022-2488/86/092217-11$02.50

of the fundamental matrix representation $D(1,0)$ of g_2 (both complex and real). All maximal subalgebras are found and the reducible ones are identified as algebras of matrices leaving different types of subspaces of $\mathbb{C}^7$ or $\mathbb{R}^7$ invariant. The main results are presented in three tables, to be further used in Paper II of this series. Throughout the article, we make use of the Chevalley basis.[25,26]

II. THE LIE GROUP $G_2(\mathbb{C})$ AND ITS LIE ALGEBRA $g_2(\mathbb{C})$

The exceptional Lie algebra $g_2(\mathbb{C})$[26,27] is one of the three existing simple rank 2 complex Lie algebras. Its order is 14, its root space is of dimension 2 and hence it has two null and 12 nonzero roots. Its root diagram has the well-known form of a "star of David."[26,28] The fundamental irreducible representations of $g_2(\mathbb{C})$, namely $D(1,0)$ and $D(0,1)$, are representations of complex dimension 7 and 14, the latter one being the adjoint or regular representation.

We shall call α_1 and α_2 the two fundamental simple roots satisfying

$$(\alpha_1,\alpha_1) = 6n, \quad (\alpha_2,\alpha_2) = 2n, \quad (\alpha_1,\alpha_2) = -3n, \tag{2.1}$$

where the normalization factor $n\in R^{>0}$ can be chosen in some convenient way. The set Δ_+ of positive roots is given by

$$\Delta_+ = (\alpha_1,\alpha_2,\alpha_1+\alpha_2,\alpha_1+2\alpha_2,\alpha_1+3\alpha_2,2\alpha_1+3\alpha_2) \tag{2.2a}$$

and the set of all roots is

$$\Delta = \Delta_+ \cup (-\Delta_+). \tag{2.2b}$$

Throughout this article we shall make use of a particularly convenient basis for g_2, namely the Chevalley basis.[25,26,29] This is an integral basis, in the sense that all structure constants are integers. It is hence highly advantageous for numerical calculations (on computers or otherwise, see, e.g., Ref. 30), for studies involving discrete subgroups or elements of finite order,[31] and also for generalizations to infinite-dimensional Lie algebras of the Kac–Moody type[32,33] (see Refs. 34 and 35).

A Chevalley basis exists for every semisimple Lie algebra L over an algebraically closed field of characteristic zero.[29] Denoting the root system Δ and the roots α_i, we can write the Chevalley basis as

$$\{h_j, e_\alpha,\ j=1,\ldots,l,\ \alpha_i\in\Delta\}. \tag{2.3a}$$

The commutation relations are

(i) $[h_i,h_j]=0, \quad 0\leqslant i, \quad j\leqslant l$;

(ii) $[h_i,e_\alpha] = 2(\alpha,\alpha_i)/(\alpha_i,\alpha_i)e_\alpha, \quad 1\leqslant i\leqslant l, \quad \alpha\in\Delta$;

(iii) $[e_\alpha,e_{-\alpha}] = h_\alpha$;

(iv) $[e_\alpha,e_\beta] = \begin{cases} \pm(r+1)e_{\alpha+\beta}, & \text{if } \alpha+\beta\in\Delta, \\ 0, & \text{if } \alpha+\beta\notin\Delta; \end{cases}$

$$\tag{2.3b}$$

where α and β are linearly independent roots and $\beta - r\alpha,\ldots,\beta-\alpha,\beta+\alpha,\ldots,\beta+q\alpha$ is the α-string through β [i.e., all members of the string belong to Δ, but $\beta-(r+1)\alpha$ and $\beta+(q+1)\alpha$ do not]. The signs in (iv) must be chosen in a consistent manner.

Returning to $g_2(\mathbb{C})$, we have the basis

$$\{h_{\alpha_1},h_{\alpha_2},e_{\pm\alpha_1},e_{\pm\alpha_2},e_{\pm(\alpha_1+\alpha_2)}, e_{\pm(\alpha_1+2\alpha_2)},e_{\pm(\alpha_1+3\alpha_2)},e_{\pm(2\alpha_1+3\alpha_2)}\}, \tag{2.4}$$

with the commutation relations (2.3), or more specifically

$$[h_{\alpha_1},h_{\alpha_2}]=0, \quad [h_\alpha,e_\beta] = 2\frac{(\alpha,\beta)}{(\alpha,\alpha)}e_\beta \equiv A_{\alpha\beta}e_\beta,$$
$$[e_\alpha,e_{-\alpha}]=h_\alpha, \quad [e_\alpha,e_\beta]=N_{\alpha\beta}e_{\alpha+\beta},$$
$$\alpha,\beta\in\Delta, \quad h_\alpha\in\mathbb{H}. \tag{2.5}$$

Here $\mathbb{H}$ is the Cartan subalgebra and the values of N_α, in agreement with (2.3), are given in Table I. For $(\alpha,\beta)=(\alpha_1,\alpha_2)$ (the fundamental roots) A is the Cartan matrix, for $g_2(C)$ equal to

$$A = \begin{pmatrix} 2 & -1 \\ -3 & 2 \end{pmatrix}. \tag{2.6}$$

The signs in Table I are chosen in such a manner that an automorphism of the g_2 root diagram extends in a simple manner[26] to an automorphism of the group $G_2(K)$, where K is a perfect field of characteristic 3.

The Cartan–Weyl basis used in much of the literature[8,22,36] differs from the Chevalley basis only by a change of normalization. Thus, let us denote the basis used, e.g., by Cornwell as[22]

$$\{\tilde h_{\alpha_1},\tilde h_{\alpha_2},X_{\pm\alpha_1},X_{\pm\alpha_2},X_{\pm(\alpha_1+\alpha_2)}, X_{\pm(\alpha_1+2\alpha_2)},X_{\pm(\alpha_1+3\alpha_2)},X_{\pm(2\alpha_1+3\alpha_2)}\}. \tag{2.7}$$

TABLE I. Values of $N_{\alpha\beta}$ for $g_2(\mathbb{C})$ in the Chevalley basis.

$N_{\alpha\beta}$	α_1	α_2	$\alpha_1+\alpha_2$	$\alpha_1+2\alpha_2$	$\alpha_1+3\alpha_2$	$2\alpha_1+3\alpha_2$	$-\alpha_1$	$-\alpha_2$	$-(\alpha_1+\alpha_2)$	$-(\alpha_1+2\alpha_2)$	$-(\alpha_1+3\alpha_2)$	$-(2\alpha_1+3\alpha_2)$
α_1	0	1	0	0	1	0	*	0	−1	0	0	−1
α_2	−1	0	−2	3	0	0	0	*	3	2	−1	0
$\alpha_1+\alpha_2$	0	2	0	3	0	0	−1	3	*	−2	0	−1
$\alpha_1+2\alpha_2$	0	−3	−3	0	0	0	0	2	−2	*	1	1
$\alpha_1+3\alpha_2$	−1	0	0	0	0	0	0	−1	0	1	*	1
$2\alpha_1+3\alpha_2$	0	0	0	0	0	0	−1	0	−1	1	1	*
$-\alpha_1$	*	0	1	0	0	1	0	−1	0	0	−1	0
$-\alpha_2$	0	*	−3	−2	1	0	1	0	2	−3	0	0
$-(\alpha_1+\alpha_2)$	1	−3	*	2	0	1	0	−2	0	−3	0	0
$-(\alpha_1+2\alpha_2)$	0	−2	2	*	−1	−1	0	3	3	0	0	0
$-(\alpha_1+3\alpha_2)$	0	1	0	−1	*	−1	1	0	0	0	0	0
$-(2\alpha_1+3\alpha_2)$	1	0	1	−1	−1	*	0	0	0	0	0	0

We then have

$$\tilde{h}_\alpha = \tfrac{1}{8}h_{\alpha_1}, \quad \tilde{h}_{\alpha_2} = \tfrac{1}{24}h_{\alpha_2}, \quad X_{\pm\alpha} = C_\alpha e_{\pm\alpha}, \tag{2.8}$$

with

$$C_{\alpha_1} = -C_{\alpha_1+3\alpha_2} = C_{2\alpha_1+3\alpha_2} = (1/\sqrt{8}),$$
$$C_{\alpha_2} = -C_{\alpha_1+\alpha_2} = C_{\alpha_1+2\alpha_2} = (1/\sqrt{24}). \tag{2.9}$$

The group $G_2(\mathbb{C})$ is isomorphic to the group of orthogonal transformations acting on a complex seven-dimensional vector space and leaving a third-order antisymmetric tensor T invariant.[27,29] Explicitly we can realize the $D(1,0)$ representation of the group $G_2(\mathbb{C})$ by the matrices $g\in\mathbb{C}^{7\times 7}$ satisfying

$$g_{ab}g_{ac} = \delta_{bc}, \tag{2.10}$$
$$g_{ab}T_{bcd} = T_{aef}g_{ec}g_{fd}, \tag{2.11}$$

where the nonzero components of the completely antisymmetric tensor T can be characterized by the values

$$T_{127} = T_{154} = T_{163} = T_{235} = T_{264} = T_{374} = T_{576} = 1. \tag{2.12}$$

The invariance of the tensor T is equivalent to the invariance of a "vector product" in seven dimensions[37,38]:

$$g(\mathbf{y}\times\mathbf{z}) = g\mathbf{y}\times g\mathbf{z}, \tag{2.13}$$

where (in a specific representation) we have

$$(y\times z)_a = \begin{pmatrix} y_{a-3} & z_{a-3} \\ y_{a-2} & z_{a-2}\end{pmatrix} + \begin{pmatrix} y_{a+2} & z_{a+2} \\ y_{a-1} & z_{a-1}\end{pmatrix} + \begin{pmatrix} y_{a+1} & z_{a+1} \\ y_{a+3} & z_{a+3}\end{pmatrix}, \quad a = 1,\ldots,7, \tag{2.14}$$

and all indices are defined mod 7. The tensor $T = \{T_{abc}\}$ is a fully antisymmetric cubic invariant,[14] satisfying "alternativity" relations.[39] An explicit realization of T can be given in terms of octonions (Cayley algebra).[13] We shall make extensive use of this tensor in Paper II of this series, when determining the nonlinear differential equations associated with G_2.

All maximal subalgebras of the complex and real forms of g_2 are determined in Sec. IV. In particular, $g_2(\mathbb{C})$ has five mutually nonisomorphic maximal reductive subalgebras. In addition to the compact real form $g_2^c(\mathbb{R})$ and the noncompact one $g_2^{\mathrm{NC}}(\mathbb{R})$ we have a class of simple sl(3,C) subalgebras, represented in the Chevalley basis by

$$\{h_{\alpha_1},h_{\alpha_2},e_{\pm\alpha_1},e_{\pm(\alpha_1+3\alpha_2)},e_{\pm(2\alpha_1+3\alpha_2)}\}. \tag{2.15}$$

A fourth class of simple Lie subalgebras is represented by the sl(2,C) algebra:

$$\{h_{\alpha_1}+h_{\alpha_2},e_{\alpha_1}+e_{\alpha_2},e_{-\alpha_1}+e_{-\alpha_2}\}. \tag{2.16}$$

Finally a class of semisimple maximal subalgebras is represented by the sl(2,C) $\oplus$ sl(2,C) subalgebra

$$\{h_{\alpha_1},e_{\alpha_1},e_{-\alpha_1}\}\oplus\{h_{\alpha_1+2\alpha_2},e_{\alpha_1+2\alpha_2},e_{-(\alpha_1+2\alpha_2)}\}. \tag{2.17}$$

The remaining maximal subalgebras of $g_2(\mathbb{C})$ are maximal parabolic subalgebras, i.e., they contain the Borel subalgebra (the maximal solvable subalgebra). For $g_2(\mathbb{C})$ the Borel subalgebra is of dimension 8 and is unique up to conjugacy (for any complex simple Lie algebra). It can be chosen to be

$$B = \{h_{\alpha_1},h_{\alpha_2},e_{\alpha_1},e_{\alpha_2},e_{\alpha_1+\alpha_2},e_{\alpha_1+2\alpha_2},e_{\alpha_1+3\alpha_2},e_{2\alpha_1+3\alpha_2}\}. \tag{2.18}$$

The algebra $g_2(\mathbb{C})$ contains two mutually nonisomorphic classes of maximal parabolic subalgebras, the so-called standard parabolic subalgebras,[22,29] given, e.g., by

$$P_{\alpha_1}(\mathbb{C}) = \{h_{\alpha_1},h_{\alpha_2},e_{\pm\alpha_1},e_{\alpha_2},e_{\alpha_1+\alpha_2}, e_{\alpha_1+2\alpha_2},e_{\alpha_1+3\alpha_2},e_{2\alpha_1+3\alpha_2}\} \tag{2.19}$$

and

$$P_{\alpha_2}(\mathbb{C}) = \{h_{\alpha_1},h_{\alpha_2},e_{\alpha_1},e_{\pm\alpha_2},e_{\alpha_1+\alpha_2}, e_{\alpha_1+2\alpha_2},e_{\alpha_1+3\alpha_2}e_{2\alpha_1+3\alpha_2}\}. \tag{2.20}$$

Notice that we have

$$B\sim P_{\alpha_1}(\mathbb{C})\cap P_{\alpha_2}(\mathbb{C}). \tag{2.21}$$

III. REAL FORMS OF g_2 AND SOME OF THEIR BASIC PROPERTIES

Let us first consider an arbitrary complex simple Lie algebra $L(\mathbb{C})$ with Chevalley basis (2.3). To this Lie algebra, we can always associate its compact real form $L^c(\mathbb{R})$ with basis[36,40,41]

$$L^c(\mathbb{R}) = \{ih_{\alpha_j},\ \epsilon_\alpha = e_\alpha - e_{-\alpha},\ \ \eta_\alpha = i(e_\alpha + e_{-\alpha}),\ j = 1,\ldots,l,\ \alpha\in\Delta_+\}. \tag{3.1}$$

The noncompact real forms $L^{\mathrm{NC}}(\mathbb{R})$ are then obtained from $L^c(\mathbb{R})$ through chief involutive automorphisms,[37] defined with respect to the Cartan subalgebra $\mathbb{H}$ of the complexification $L(\mathbb{C})$ of $L^{\mathrm{NC}}(\mathbb{R})$.

The Cartan decomposition for a noncompact real form $L^{\mathrm{NC}}(\mathbb{R})$ takes the form

$$L^{\mathrm{NC}}(\mathbb{R}) = K + P, \tag{3.2}$$

where K is a maximal compact subalgebra of L (unique up to conjugacy) satisfying

$$K = \{a\in L\,|Za = a\}. \tag{3.3}$$

The subspace P satisfies

$$P = \{a\in L\,|Za = -a\}. \tag{3.4}$$

For the present purposes we can restrict ourselves to chief inner automorphisms[37,21] and we have

$$Z = \exp(adh), \quad h\in\mathbb{H}. \tag{3.5}$$

This automorphism is diagonal with respect to the canonical basis (3.1) of $L^c(\mathbb{R})$. The basis elements ih_{α_j} $(j = 1,\ldots,l)$ correspond to the eigenvalue $+1$; ϵ_α and η_α correspond to the eigenvalue $\exp\{\alpha(h)\} = \pm 1$, where $\alpha(h) = B(h,h_\alpha)$ [$B(x,y)$ is the Killing form of $L(\mathbb{C})$]. We hence obtain

$$K = \{ih_{\alpha_j},\epsilon_\alpha,\eta_\alpha,\ i = 1,\ldots,l\ (\alpha|\exp\alpha(h) = 1)\},$$
$$P = \{i\epsilon_\alpha, -i\eta_\alpha\ (\alpha|\exp\alpha(h) = -1)\}. \tag{3.6}$$

In the case of the exceptional Lie algebra g_2 there exist two nonisomorphic real forms,[37,38] the compact form $g_2^{\mathrm{C}}(\mathbb{R})$ with the character -14 and the noncompact one $g_2^{\mathrm{NC}}(\mathbb{R})$ with the character $+2$. Correspondingly, two nonequivalent chief inner automorphisms exist in this case. The first is given by the choice

$$\exp \alpha_1(h) = \exp \alpha_2(h) = 1, \tag{3.7}$$

i.e., Z is the identity and we obtain $g_2^{\mathrm{C}}(\mathbf{R})$ itself:

$$g_2^{\mathrm{C}}(\mathbf{R}) \sim \{ih_{\alpha_1}, ih_{\alpha_2}, \epsilon_{\alpha_1}, \eta_{\alpha_1}, \epsilon_{\alpha_2}, \eta_{\alpha_2}, \epsilon_{\alpha_1+\alpha_2}, \eta_{\alpha_1+\alpha_2}, \epsilon_{\alpha_1+2\alpha_2}, \eta_{\alpha_1+2\alpha_2}, \epsilon_{\alpha_1+3\alpha_2}, \eta_{\alpha_1+3\alpha_2}, \epsilon_{2\alpha_1+3\alpha_2}, \eta_{2\alpha_1+3\alpha_2}\}. \tag{3.8}$$

The second choice of automorphism is given by

$$\exp \alpha_1(h) = 1, \quad \exp \alpha_2(h) = -1, \tag{3.9}$$

and provides us with a basis for the noncompact form $g_2^{\mathrm{NC}}(\mathbf{R})$. In (3.2) we have, in this case,

$$g_2^{\mathrm{NC}}(\mathbf{R}) = \{K, P\}, \tag{3.10a}$$

$$K = \{ih_{\alpha_1}, ih_{\alpha_2}, \epsilon_{\alpha_1}, \eta_{\alpha_1}, \epsilon_{\alpha_1+2\alpha_2}, \eta_{\alpha_1+2\alpha_2}\}, \tag{3.10b}$$

$$P = \{i\epsilon_{\alpha_2}, -i\eta_{\alpha_2}, i\epsilon_{\alpha_1+\alpha_2}, -i\eta_{\alpha_1+\alpha_2}, i\epsilon_{\alpha_1+3\alpha_2}, -i\eta_{\alpha_1+3\alpha_2}, i\epsilon_{2\alpha_1+3\alpha_2}, -i\eta_{2\alpha_1+3\alpha_2}\}. \tag{3.10c}$$

An important concept for a noncompact Lie algebra is that of the Iwasawa[23] decomposition:

$$L = K + A + N, \tag{3.11}$$

where K is a maximal compact subalgebra, A is a maximal Abelian subalgebra of P, with $\dim A = m < l$, and N is a nilpotent subalgebra of L. Cornwell has given a direct prescription[22] for calculating the Iwasawa decomposition and has applied it to the pseudo-orthogonal Lie algebras so(3,1), so(4,1), so(3,2), and so(4,2). We applied Cornwell's method to $g_2^{\mathrm{NC}}(\mathbf{R})$. Dropping all details, we simply give the results. If we choose the maximal compact subalgebra K of $g_2^{\mathrm{NC}}(\mathbf{R})$ in the form (3.10b) and use the basics (3.10), we find

$$A = \{i\epsilon_{\alpha_2}, i\epsilon_{2\alpha_1+3\alpha_2}\} \tag{3.12}$$

and

$$N = \{\epsilon_{\alpha_1} + i\epsilon_{\alpha_1+3\alpha_2}, \epsilon_{\alpha_1+2\alpha_2} - i\epsilon_{\alpha_1+\alpha_2}, \eta_{\alpha_1} - i\eta_{\alpha_1+3\alpha_2}, \eta_{\alpha_1+2\alpha_2} + i\eta_{\alpha_1+\alpha_2}, i\eta_{\alpha_2} + ih_{\alpha_2}, -i\eta_{2\alpha_1+3\alpha_2} + i(2h_{\alpha_1} + h_{\alpha_2})\}. \tag{3.13}$$

Furthermore

$$p_m = \{A + N\} \tag{3.14}$$

is a "minimal parabolic subalgebra" of $g_2^{\mathrm{NC}}(\mathbf{R})$; it is a maximal solvable subalgebra[42] and its complexification is the Borel subalgebra of $g_2(\mathbf{C})$. Following Cornwell's prescription, we could use the Iwasawa decomposition to obtain the two "standard parabolic subalgebras" of $g_2^{\mathrm{NC}}(\mathbf{R})$. Again, we only present the results and moreover, we shall use a different basis and choose a different (equivalent) realization of the maximal compact subalgebra K.

Indeed, consider the Chevalley basis (2.4) of $g_2(\mathbf{C})$, however, consider it over the field of real numbers $\mathbf{R}$, rather than over $\mathbf{C}$. In this case (2.4) immediately provides a basic for $g_2^{\mathrm{NC}}(\mathbf{R})$. The maximal compact subalgebra is given as

$$K = \{e_{\alpha_1} - e_{-\alpha_1}, e_{\alpha_2} - e_{-\alpha_2}, e_{\alpha_1+\alpha_2} - e_{-\alpha_1-\alpha_2}, e_{\alpha_1+2\alpha_2} - e_{-\alpha_1-2\alpha_2}, e_{\alpha_1+3\alpha_2} - e_{-\alpha_1-3\alpha_2}, e_{2\alpha_1+3\alpha_2} - e_{-2\alpha_1-3\alpha_2}\}, \tag{3.15}$$

the Abelian algebra A of (3.11) is

$$A = \{h_{\alpha_1}, h_{\alpha_2}\} \tag{3.16}$$

and the nilpotent algebra N is

$$N = \{e_{\alpha_1}, e_{\alpha_2}, e_{\alpha_1+\alpha_2}, e_{\alpha_1+2\alpha_2}, e_{\alpha_1+3\alpha_2}, e_{2\alpha_1+3\alpha_2}\}. \tag{3.17}$$

The two nonisomorphic "standard parabolic subalgebras" (that are also maximal parabolic subalgebras) are

$$P_{\alpha_1}(\mathbf{R}) = \{h_{\alpha_1}, h_{\alpha_2}, e_{-\alpha_1}, e_{\alpha_1}, e_{\alpha_2}, e_{\alpha_1+\alpha_2}, e_{\alpha_1+2\alpha_2}, e_{\alpha_1+3\alpha_2}, e_{2\alpha_1+3\alpha_2}\}. \tag{3.18}$$

and

$$P_{\alpha_2}(\mathbf{R}) = \{h_{\alpha_1}, h_{\alpha_2}, e_{-\alpha_2}, e_{\alpha_1}, e_{\alpha_2}, e_{\alpha_1+\alpha_2}, e_{\alpha_1+2\alpha_2}, e_{\alpha_1+3\alpha_2}, e_{2\alpha_1+3\alpha_2}\}. \tag{3.19}$$

As in the complex case we have

$$P_m = P_{\alpha_1}(\mathbf{R}) \cap P_{\alpha_2}(\mathbf{R}), \tag{3.20}$$

where P_m is the "minimal parabolic subalgebra."

Notice that the complexification of (3.18), (3.19), and (3.20) are (2.19), (2.20), and (2.18), respectively.

All maximal subalgebras of $g_2(\mathbf{C})$, $g_2^{\mathrm{C}}(\mathbf{R})$, and $g_2^{\mathrm{NC}}(\mathbf{R})$ are derived in the following section.

IV. THE MATRIX REPRESENTATION $D(1,0)$ AND THE MAXIMAL SUBALGEBRAS OF g_2

The fundamental representation $D(1,0)$ of g_2 is of dimension 7. This is the lowest-dimensional faithful representation of g_2 and as such, it is particularly convenient for visualizing subalgebras of $g_2(\mathbf{C})$ and its two real forms.

For $g_2(\mathbf{C})$ this representation can be viewed as a restriction of the defining representation of o(7,C) (the 21-dimensional classical simple Lie algebra B_3).[27,29] For $g_2^{\mathrm{C}}(\mathbf{R})$ and $g_2^{\mathrm{NC}}(\mathbf{R})$ the representation $D(1,0)$ is a restriction of the defining representation of the real algebra o(7), or o(4,3), respectively.

A. Matrix realizations of the complex and real forms of g_2

Cartan[27] has proposed a specific realization of the algebra $g_2(\mathbf{C})$ when this algebra acts on a seven-dimensional vector space V_7 characterized by coordinates x_j, z, y_i $(i = 1,2,3)$:

$$V_7^T = \{x_1, x_2, x_3, z, y_1, y_2, y_3\}. \tag{4.1}$$

The infinitesimal operators are given as vector fields

$$X_{ii} = -x_i \frac{\partial}{\partial x_i} + y_i \frac{\partial}{\partial y_i} + \frac{1}{3}\left(x_k \frac{\partial}{\partial x_k} - y_k \frac{\partial}{\partial y_k}\right)$$

(no sum over i),

$$X_{i0} = -2z \frac{\partial}{\partial x_i} + y_i \frac{\partial}{\partial z} + \frac{1}{2}\epsilon_{ijk}\left(x_j \frac{\partial}{\partial y_k} - x_k \frac{\partial}{\partial y_j}\right),$$

$$X_{0i} = 2z \frac{\partial}{\partial y_i} + y_i \frac{\partial}{\partial z} + \frac{1}{2}\epsilon_{ijk}\left(y_j \frac{\partial}{\partial x_k} - y_k \frac{\partial}{\partial x_j}\right),$$

$$X_{ij} = -x_j \frac{\partial}{\partial x_i} + y_i \frac{\partial}{\partial y_j} \quad (i \neq j). \tag{4.2}$$

Note that we have

$$X_{11}+X_{22}+X_{33}=0, \tag{4.3}$$

so that (4.2) only represents 14 operators.

The commutation relations of the $g_2(\mathbb{C})$ Lie algebra in this basis are

$$\begin{aligned}
&[X_{ii},X_{i0}]=\tfrac{2}{3}X_{i0}, \quad [X_{ii},X_{k0}]=-\tfrac{1}{3}X_{k0} \quad (i\neq k),\\
&[X_{ii},X_{0i}]=-\tfrac{2}{3}X_{0i}, \quad [X_{ii},X_{0k}]=\tfrac{1}{3}X_{0k} \quad (i\neq k),\\
&[X_{i0},X_{j0}]=2\epsilon_{ijk}X_{0k}, \quad [X_{0i},X_{0j}]=-2\epsilon_{ijk}X_{k0},\\
&[X_{0i},X_{jk}]=\delta_{ij}X_{0k}, \quad [X_{ik},X_{j0}]=\delta_{kj}X_{i0},\\
&[X_{i0},X_{0j}]=3X_{ij}, \quad [X_{ij},X_{lm}]=\delta_{jl}X_{im}-\delta_{mi}X_{lj}.
\end{aligned} \tag{4.4}$$

We shall eliminate one of three generators X_{ii} [see (4.3)] and use the following basis:

$$\{X_1=X_{11}-X_{22}, X_2=-3X_{11}, X_{i0},X_{0i},X_{ij}\},$$
$$i\neq j, \quad i,j=1,2,3. \tag{4.5}$$

Such a basis is related in a simple way to the Chevalley basis (2.4) of Sec. II. Indeed, we have

$$\begin{aligned}
&X_1=h_{\alpha_1}, \quad X_2=h_{\alpha_2},\\
&X_{10}=e_{-\alpha_2}, \quad X_{20}=-e_{-(\alpha_1+\alpha_2)}, \quad X_{30}=-e_{\alpha_1+2\alpha_2},\\
&X_{01}=e_{\alpha_2}, \quad X_{02}=-e_{\alpha_1+\alpha_2}, \quad X_{03}=-e_{-(\alpha_1+2\alpha_2)},\\
&X_{12}=e_{\alpha_1}, \quad X_{23}=-e_{-(2\alpha_1+3\alpha_2)}, \quad X_{31}=e_{\alpha_1+3\alpha_2},\\
&X_{21}=e_{-\alpha_1}, \quad X_{32}=-e_{2\alpha_1+3\alpha_2}, \quad X_{13}=e_{-(\alpha_1+3\alpha_2)}.
\end{aligned} \tag{4.6}$$

It is easy to pass from the realization (4.2) of the $g_2(\mathbb{C})$ generators as differential operators to a seven-dimensional matrix representation. We shall need several different matrix realizations for different applications. The algebra $g_2(\mathbb{C})$ will be viewed as a subalgebra of $o(7,\mathbb{C})$. A matrix $M\in o(7,\mathbb{C})$ satisfies

$$KM+M^TK=0, \quad K=K^T, \quad \det K\neq 0, \tag{4.7}$$

where the superscript T denotes transposition and K is a fixed symmetric nonsingular matrix. We can pass from a realization corresponding to a given metric tensor K to a different realization with metric K' by putting

$$K=S^TK'S, \quad M=S^{-1}M'S. \tag{4.8}$$

One convenient choice is due to Cartan[27]:

$$K_C=\begin{pmatrix} 0 & 0 & I_3/2\\ 0 & 1 & 0\\ I_3/2 & 0 & 0\end{pmatrix}, \tag{4.9}$$

where I_n is the n-dimensional identity matrix. Then we have

$$M_C=\begin{pmatrix} A & 2b & C\\ d^T & 0 & -b^T\\ D & -2d & -A^T\end{pmatrix},$$
$$A,C,D\in\mathbb{C}^{3\times 3}, \quad C^T+C=0,$$
$$D^T+D=0, \quad b,d\in\mathbb{C}^{3\times 1}. \tag{4.10a}$$

For M_C to lie in the $g_2(\mathbb{C})$ subalgebra of $o(7,\mathbb{C})$ we must impose $\operatorname{Tr} A=0$ and relate C and D to d and b, respectively:

$$A=\begin{pmatrix} a_1-2a_2 & a_{12} & a_{13}\\ a_{21} & -a_1+a_2 & a_{23}\\ a_{31} & a_{32} & a_2\end{pmatrix},$$

$$C=\begin{pmatrix} 0 & -a_{03} & a_{02}\\ a_{03} & 0 & -a_{01}\\ -a_{02} & a_{01} & 0\end{pmatrix},$$

$$D=\begin{pmatrix} 0 & a_{30} & -a_{20}\\ -a_{30} & 0 & a_{10}\\ a_{20} & -a_{10} & 0\end{pmatrix},$$

$$b=\begin{pmatrix} a_{10}\\ a_{20}\\ a_{30}\end{pmatrix}, \quad d=\begin{pmatrix} a_{01}\\ a_{02}\\ a_{03}\end{pmatrix}. \tag{4.10b}$$

The matrix of the form M_C representing the generator $X_{\mu\nu}$ $(\mu,\nu=0,1,2,3)$ or X_a $(a=1,2)$ of (4.5) is obtained by putting the corresponding $a_{\mu\nu}=1$ (or $a_i=1$) in (4.10), and all other entries $a_{\mu'\nu'}=0$, $a_{i'}=0$.

Note that in this realization the orthogonal group $O(7,\mathbb{C})$ leaves the quadratic form

$$Q=X^TK_CX=z^2+(\mathbf{x},\mathbf{y}) \tag{4.11}$$

invariant, where

$$X^T=(\mathbf{x},z,\mathbf{y}), \quad \mathbf{x},\mathbf{y}\in\mathbb{C}^3, \quad z\in\mathbb{C}.$$

In terms of the more usual diagonal metric, I_7, we have

$$K_C=S^TI_7S, \quad M=SM_CS^{-1}, \tag{4.12}$$

with

$$S=\begin{pmatrix} I_3/2 & 0 & I_3/2\\ 0 & 1 & 0\\ -iI_3/2 & 0 & iI_3/2\end{pmatrix}. \tag{4.13}$$

Explicitly, we have $M+M^T=0$ and hence

$$M=\begin{pmatrix} R & m & V\\ -m^T & 0 & n^T\\ -V^T & -n & U\end{pmatrix},$$
$$R,U,V\in\mathbb{C}^{3\times 3}, \quad R^T+R=0,$$
$$U^T+U=0, \quad m,n\in\mathbb{C}^{3\times 1}, \tag{4.14}$$

where

$$\begin{aligned}
&2R=(A-A^T+C+D), \quad 2U=(A-A^T-C-D),\\
&2V=i(A+A^T-C+D), \quad m=b-d, \quad n=i(b+d).
\end{aligned} \tag{4.15}$$

In this realization the completely antisymmetric tensor T of (2.11) has the components (2.12) and in Lie algebraic terms the invariance condition (2.11) translates into

$$M_{ab}T_{bcd}=[T_a,M]_{cd}=T_{acf}M_{fd}-M_{ce}T_{aed}. \tag{4.16}$$

Two further realizations will be needed below. For the first we put

$$K_C=S'^TI_{4,3}S', \quad I_{4,3}=\begin{pmatrix} I_4 & \\ & -I_3\end{pmatrix}, \tag{4.17}$$

$$S'=\begin{pmatrix} I_3/2 & 0 & I_3/2\\ 0 & 1 & 0\\ -I_3/2 & 0 & I_3/2\end{pmatrix}$$

and obtain

$$M' = \begin{pmatrix} R & m & iV \\ -m^T & 0 & in^T \\ iV^T & in & U \end{pmatrix},$$

$$R^T + R = 0, \quad U^T + U = 0, \tag{4.18}$$

with R, U, V, m, and n as in (4.15).

Finally, let J_7 be the antidiagonal matrix

$$J_7 = \{\delta_{a,8-b}, \quad a,b = 1,...,7\}. \tag{4.19}$$

We have $J_7 = S''^T K_C S''$ with

$$S'' = \begin{bmatrix} 0 & 0 & \sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{2} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sqrt{2} & 0 \\ \sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{4.20}$$

and

$$M'' = S''^{-1} M_C S'' = \begin{bmatrix} a_2 & -a_{10} & a_{20} & -\sqrt{2}a_{03} & -a_{13} & -a_{23} & 0 \\ -a_{01} & -a_1 + a_2 & a_{21} & \sqrt{2}a_{20} & a_{03} & 0 & a_{23} \\ a_{02} & a_{12} & a_1 - 2a_2 & \sqrt{2}a_{10} & 0 & -a_{03} & a_{13} \\ -\sqrt{2}a_{30} & \sqrt{2}a_{02} & \sqrt{2}a_{01} & 0 & -\sqrt{2}a_{10} & -\sqrt{2}a_{20} & \sqrt{2}a_{03} \\ -a_{31} & a_{30} & 0 & -\sqrt{2}a_{01} & -a_1 + 2a_2 & -a_{21} & -a_{20} \\ -a_{32} & 0 & -a_{30} & -\sqrt{2}a_{02} & -a_{12} & a_1 - a_2 & a_{10} \\ 0 & a_{32} & a_{31} & \sqrt{2}a_{30} & -a_{02} & a_{01} & a_2 \end{bmatrix}. \tag{4.21}$$

B. The maximal subalgebras of $g_2(\mathbb{C})$

The maximal subalgebras of a semisimple Lie algebra can be embedded reducibly or irreducibly in a given finite-dimensional representation.[43] The reducibly embedded ones leave some vector subspace invariant, the irreducibly embedded ones do not. Irreducibly embedded subalgebras are always reductive (semisimple, or the direct sum of a semisimple Lie algebra with an Abelian one.[43] The semisimple subalgebras of the simple complex Lie algebras were classified by Dynkin,[44] those of the real simple Lie algebras by Cornwell.[45]

The algebra $g_2(\mathbb{C})$ has three irreducibly embedded subalgebras (up to conjugacy), all of them simple. They are as follows.

(1) $g_2^{\mathbb{C}}(\mathbb{R})$, the maximal compact subalgebra of $g_2(\mathbb{C})$. It is best obtained in the realization (4.14) by restricting all entries in R, U, V, and m and n to be real. We then have

$$g_2^{\mathbb{C}}(\mathbb{R}) \sim g_2(\mathbb{C}) \cap o(7), \tag{4.22}$$

where both $g_2(\mathbb{C})$ and $o(7)$ are realized using the metric $K = I_7$.

The identification (3.8) of $g_2^{\mathbb{C}}(\mathbb{R})$ in Sec. III in terms of the Chevalley basis is equivalent to that given by (4.22) [or (4.14) with real entries], but does not coincide with this realization. Indeed, the choice (3.8) corresponds to

$$g_2^{\mathbb{C}}(\mathbb{R}) \sim g_2(\mathbb{C}) \cap su(7), \tag{4.23}$$

where $g_2(\mathbb{C})$ is taken in the realization (4.10) [with metric (4.9)] and $su(7)$ is realized by matrices $X \in \mathbb{C}^{7\times 7}$ satisfying

$$\tilde{I}X + X^\dagger \tilde{I} = 0, \quad \tilde{I} = \begin{pmatrix} I_3 & & \\ & 2 & \\ & & I_3 \end{pmatrix} \tag{4.24}$$

(the superscript † denotes Hermitian conjugation).

(2) $g_2^{\mathrm{NC}}(\mathbb{R})$, the noncompact real form of g_2. This subalgebra can be easily obtained in the realization (4.18), corresponding to the metric $I_{4,3}$ of (4.17), by requiring that R, m, and U be real and V and n be pure imaginary. We then have

$$g_2^{\mathrm{NC}}(\mathbb{R}) \sim g_2(\mathbb{C}) \cap O(4,3), \tag{4.25}$$

where both $g_2(\mathbb{C})$ and $O(4,3)$ are realized using the metric $I_{4,3}$.

The realization (3.10) of Sec. III is equivalent to this one, though it does not coincide with it. The choice (3.10) corresponds to the intersection

$$g_2^{\mathrm{NC}}(\mathbb{R}) \sim g_2(\mathbb{C}) \cap su(4,3), \tag{4.26}$$

where $g_2(\mathbb{C})$ is realized using the metric $K = K_C$ of (4.9) and $su(4,3)$ is realized by the matrices $X \in \mathbb{C}^{7\times 7}$, satisfying

$$\tilde{\tilde{I}}X + X^+ \tilde{\tilde{I}} = 0,$$

$$\tilde{\tilde{I}} = \begin{pmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & -1 & & & & \\ & & & -2 & & & \\ & & & & 1 & & \\ & & & & & 1 & \\ & & & & & & -1 \end{pmatrix}. \tag{4.27}$$

(3) The algebra $sl(2,\mathbb{C})$, already given in (2.16) is realized, e.g., by the matrices

$$\{X_{12} + X_{01}, X_{21} + X_{10}, X_1 + X_2\} \tag{4.28}$$

in the realization (4.9) and (4.10).

All other maximal subalgebras of $g_2(\mathbb{C})$ are embedded reducibly in the representation $D(1,0)$. As such, they must leave a vector subspace of $\mathbb{C}^7$ invariant. The metric K provides us with an invariant vector product $x^T K y$ in $\mathbb{C}^7$. A vector $x \in \mathbb{C}^7$ can thus be either nonisotropic [$x^T K x \neq 0$, we denote such a vector space $(+)$], or isotropic [$x^T K x = 0$, we denote such a space (0)]. A subspace can be characterized by its dimension and by the number n_0 of isotropic vectors in an orthogonal basis ($0 < n_0 < 3$). If a degenerate space ($n_0 > 1$) is left invariant by some group G, then its isotropic subspace (of dimension n_0) is itself invariant. If a subspace V is invariant under G, then its orthogonal complement $V^\perp$ (with respect to the invariant metric) is also invariant.

In order to find all reducibly embedded maximal subalgebras of $g_2(\mathbb{C})$, we must hence find the subalgebras leaving invariant spaces of the type $(+)$, $(++)$, $(+++)$, (0), (00), and (000). Let us consider the individual cases.

(1) $(+)$. Consider the realization (4.9). With no loss of generality, we can choose the vectors space $(+)$ in the form

$$V_1^T = \{(0\,0\,0\,x\,0\,0\,0)\}. \tag{4.29}$$

Imposing

$$M_{\mathbb{C}} V \sim V,$$

we obtain $b = d = 0$ in (4.10), which implies $C = D = 0$. We obtain the algebra $\mathrm{sl}(3,\mathbb{C})$ realized as

$$\{X\} = \begin{pmatrix} A & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -A^T \end{pmatrix}, \quad A \in \mathbb{C}^{3\times 3}, \quad \mathrm{Tr}\, A = 0. \tag{4.30}$$

This actually coincides with the subalgebra (2.15). We have

$$\mathrm{sl}(3,\mathbb{C}) \sim o(6,\mathbb{C}) \cap g_2(\mathbb{C}).$$

(2) $(++)$. In this case, we use the realization (4.14) with $(K = I_7)$. With no loss of generality we choose the invariant subspace to be

$$V_2^T = (x\,0\,0\,y\,0\,0\,0). \tag{4.31}$$

Putting $MV \subseteq V$ we find a four-dimensional $\mathrm{gl}(2,\mathbb{C})$ algebra. This algebra is contained in a $\mathrm{sl}(3,\mathbb{C})$ algebra conjugate to (4.30) and is hence not maximal.

(3) $(+++)$. We again use the realization (4.14) and choose the invariant subspace in the form

$$V_3^T = (0\,0\,0\,0\,x\,y\,z). \tag{4.32}$$

Requiring that the space (4.32) be left invariant by a subalgebra of matrices of the form (4.14) leads to the subalgebra $\mathrm{sl}(2,\mathbb{C}) \oplus \mathrm{sl}(2,\mathbb{C})$, realized as

$$\{X\} = \begin{pmatrix} 0 & z_3 & -z_2 & z_1 & & & \\ -z_3 & 0 & z_1 & z_2 & & & \\ z_2 & -z_1 & 0 & z_3 & & & \\ -z_1 & -z_2 & -z_3 & 0 & & & \\ & & & & 0 & 0 & 0 \\ & & & & 0 & 0 & 0 \\ & & & & 0 & 0 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & -y_3 & y_2 & y_1 & & & \\ y_3 & 0 & -y_1 & y_2 & & & \\ -y_2 & y_1 & 0 & y_3 & & & \\ -y_1 & -y_2 & -y_3 & 0 & & & \\ & & & & 0 & -2y_3 & 2y_2 \\ & & & & 2y_3 & 0 & -2y_1 \\ & & & & -2y_2 & 2y_1 & 0 \end{pmatrix}. \tag{4.33}$$

We see that in this case we have

$$\mathrm{sl}(2,\mathbb{C}) \oplus \mathrm{sl}(2,\mathbb{C}) \sim g_2(\mathbb{C}) \cap [o(4,\mathbb{C}) \oplus o(3,\mathbb{C})]. \tag{4.34}$$

This algebra is conjugate to (2.17) (but does not coincide with it).

The remaining invariant subspaces to be considered are completely isotropic. For the classical groups invariance of an isotropic subspace leads to parabolic subgroups (and their Lie algebras[46]). We shall see that for $g_2(\mathbb{C})$ this is not always the case.

(4) The space (0). We use the realization (4.21) with $K = J_7$ (4.19). We choose the invariant subspace in the form

$$L_1^T = (0\,0\,0\,0\,0\,0\,x). \tag{4.35}$$

Requiring $M'' L^T \sim L^T$ implies

$$a_{23} = a_{13} = a_{03} = a_{20} = a_{10} = 0 \tag{4.36}$$

and we obtain the maximal parabolic subalgebra $P_{\alpha_1}(\mathbb{C})$ of (2.19).

Notice that the nine-dimensional algebra $P_{\alpha_{1+}}(\mathbb{C})$ can be interpreted as the restriction of the 16-dimensional similitude algebra[47] $\mathrm{sim}(5,\mathbb{C})$ to $g_2(\mathbb{C})$:

$$P_{\alpha_1}(\mathbb{C}) \sim \mathrm{sim}(5,\mathbb{C}) \cap g_2(\mathbb{C}), \tag{4.37}$$

where $\mathrm{sim}(5,\mathbb{C})$ is one of the maximal parabolic subalgebras of $o(7,\mathbb{C})$ [the group $\mathrm{sim}(5,\mathbb{C})$ is the group of Euclidian transformations of $\mathbb{C}^5$, extended by dilations].

(5) The space (00). We again use the realization (4.21) and require that the subspace

$$L_2^T = (0\,0\,0\,0\,0\,x\,y) \tag{4.38}$$

be left invariant. In (4.21), we then have

$$a_{23} = a_{03} = a_{13} = a_{20} = a_{21} = 0 \tag{4.39}$$

and we obtain the maximal parabolic subalgebra $P_{\alpha_2}(\mathbb{C})$ of (2.20). This nine-dimensional algebra [not isomorphic to $P_{\alpha_1}(\mathbb{C})$] can be interpreted as the restriction of the 14-dimensional "optical" subalgebra[47] $\mathrm{opt}(5,\mathbb{C})$ of $o(7,\mathbb{C})$ to $g_2(\mathbb{C})$:

$$P_{\alpha_2}(\mathbb{C}) \sim \mathrm{opt}(5,\mathbb{C}) \cap g_2(\mathbb{C}). \tag{4.40}$$

(6) The space (000). We use the realization (4.10) and choose

$$L_3^T = (0\,0\,0\,0\,x\,y\,z). \tag{4.41}$$

The condition $M_{\mathbb{C}} L_{3-} \subseteq L_3$ implies $b = 0$, $C = 0$. This would provide us with a new 15-dimensional maximal parabolic subalgebras of $o(7,\mathbb{C})$. Restricting to $g_2(\mathbb{C})$, we find that $C = 0$ implies $d = 0$ and $b = 0$ implies $D = 0$ [see (4.10)]. We do not obtain a new parabolic subalgebra of $g_2(\mathbb{C})$ but simply reobtain the maximal reductive subalgebra $\mathrm{sl}(3,\mathbb{C})$.

Thus, we find that $g_2(\mathbb{C})$ has precisely seven classes of maximal subalgebras, summarized in Table II.

C. The maximal subalgebras of $g_2^{\mathbb{C}}(\mathbb{R})$

All maximal subalgebras of $g_2^{\mathbb{C}}(\mathbb{R})$ (and of any compact Lie algebra) are reductive. We shall use the $D(1,0)$ representation, in which $g_2^{\mathbb{C}}(\mathbb{R})$ is viewed as a subalgebra of $o(7)$ and choose the metric to be given by $K = I_7$. Thus we have

TABLE II. The complex Lie algebra $g_2(\mathbb{C})$ and its maximal subalgebras.

Algebra	Complex ($d_{\mathbb{C}}$) or real ($d_{\mathbb{R}}$) dimension	Basis and matrix realization	Invariant subspace and metric used
$g_2(\mathbb{C})$	$d_{\mathbb{C}}=14$	(2.4) and (4.10), (4.14), (4.18) or (4.21) with $M_{ik}\in\mathbb{C}$	...
$g_2^{\mathrm{C}}(\mathbb{R})$	$d_{\mathbb{R}}=14$	(3.8) and (4.14) with $M_{ik}\in\mathbb{R}$	...
$g_2^{\mathrm{NC}}(\mathbb{R})$	$d_{\mathbb{R}}=14$	(3.10) and (4.10), (4.18), or (4.21) with $M_{ik}\in\mathbb{R}$	...
sl(2,C)	$d_{\mathbb{C}}=3$	(2.16) and (4.28)	...
sl(3,C)	$d_{\mathbb{C}}=8$	(2.15) and (4.30)	$V_1^T=\{000x000\}$, $K=K_{\mathrm{C}}$
sl(2,C) $\oplus$ sl(2,C)	$d_{\mathbb{C}}=6$	(2.17) and (4.33)	$V_3^T=\{0000xyz\}$, $K=I_7$
$P_{\alpha_1}(\mathrm{C})$	$d_{\mathbb{C}}=9$	(2.19) and (4.21) with(4.36)	$L_1^T=\{000000x\}$, $K=J_7$
$P_{\alpha_2}(\mathrm{C})$	$d_{\mathbb{C}}=9$	(2.20) and (4.21) with (4.39)	$L_2^T=\{00000xy\}$, $K=J_7$

the matrices M of (4.14) with all entries real. In other words we have

$$(b-d)\in\mathbb{R}^3,\quad (A-A^T),(C+D)\in\mathbb{R}^{3\times 3},\quad i(b+d)\in\mathbb{R}^3,\quad i(A+A^T),\quad i(C-D)\in\mathbb{R}^{3\times 3}. \tag{4.42}$$

The only subalgebra of $g_2^{\mathrm{C}}(\mathbb{R})$ irreducibly embedded in this representation is su(2). A basis for this "irreducible" su(2) algebra is given by

$$i(h_{\alpha_1}+h_{\alpha_2}) = \begin{pmatrix} 0 & 0 & 0 & 0 & -1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 1\\ 1 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\epsilon_{\alpha_1}+\epsilon_{\alpha_2} = \begin{pmatrix} 0 & 1 & 0 & -2 & 0 & 0 & 0\\ -1 & 0 & -1 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 0 & 0\\ 2 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & -1 & 0 & 1\\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \end{pmatrix},$$

$$\eta_{\alpha_1}+\eta_{\alpha_2} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & -1 & 0\\ 0 & 0 & 0 & 0 & -1 & 0 & -1\\ 0 & 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & -2 & 0 & 0\\ 0 & 1 & 0 & 2 & 0 & 0 & 0\\ 1 & 0 & -1 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{4.43}$$

The reducibly embedded maximal subalgebras are obtained by requiring that either a one-dimensional or three-dimensional vector space be left invariant (a two-dimensional space leads to a nonmaximal subalgebra). Since the metric is positive definite, all vectors have positive length.

A one-dimensional vector space can be chosen to be $V_1^T=(000x000)$. Its invariance implies $m=n=0$ in (4.14) and we obtain the algebra su(3) realized as

$$M=\begin{pmatrix} (A-A^T)/2 & 0 & i(A+A^T)/2\\ 0 & 0 & 0\\ -i(A+A^T)/2 & 0 & (A-A^T)/2 \end{pmatrix},$$

$$A-A^T\in\mathbb{R},\quad i(A+A^T)\in\mathbb{R}. \tag{4.44}$$

We have in this realization

$$\mathrm{su}(3)\sim g_2^{\mathrm{C}}(\mathbb{R})\cap \mathrm{O}(6). \tag{4.45}$$

In terms of the canonical basis (3.8) we have

$$\mathrm{su}(3)\sim\{ih_{\alpha_1},ih_{\alpha_2},\epsilon_{\alpha_1},\eta_{\alpha_1},\epsilon_{\alpha_1+3\alpha_2},\eta_{\alpha_1+3\alpha_2},\epsilon_{2\alpha_1+3\alpha_2},\eta_{2\alpha_1+3\alpha_2}\}. \tag{4.46}$$

Finally, a three-dimensional invariant subspace can be chosen to be V_3 of (4.32). Its invariance leads to an su(2) $\oplus$ su(2) subalgebra of the form (4.33) (with real entries). We have

$$\mathrm{su}(2)\oplus\mathrm{su}(2)\sim g_2^{\mathrm{C}}(\mathbb{R})\cap[\mathrm{o}(4)\oplus\mathrm{o}(3)]. \tag{4.47}$$

Equivalently, the su(2) $\oplus$ su(2) subalgebra in the canonical basis can be identified as

$$\{ih_{\alpha_1},\epsilon_{\alpha_1},\eta_{\alpha_1}\}\oplus\{i(h_{\alpha_1}+2h_{\alpha_2}),\epsilon_{\alpha_1+2\alpha_2},\eta_{\alpha_1+2\alpha_2}\}. \tag{4.48}$$

The results on the maximal subalgebras of $g_2^{\mathrm{C}}(\mathbb{R})$ are summarized in Table III.

D. The maximal subalgebras of $g_2^{\mathrm{NC}}(\mathbb{R})$

Similarly as $g_2^{\mathrm{C}}(\mathbb{R})$ the noncompact real form $g_2^{\mathrm{NC}}(\mathbb{R})$ has just one maximal subalgebra that is irreducibly embedded in the seven-dimensional fundamental representation. In this case, the subalgebra is su(1,1). To visualize it, let us take the "antidiagonal" realization of o(4,3), i.e., the metric (4.9) and the realization (4.10) with all entries real.The algebra su(1,1) is given by

$$h_{\alpha_1}+h_{\alpha_2}=X_1+X_2,\quad e_{\alpha_1}+e_{\alpha_2}=X_{12}+X_{01},\quad e_{-\alpha_1}+e_{-\alpha_2}=X_{21}+X_{10}, \tag{4.49}$$

i.e., its basis coincides with the basis (4.28) for sl(2,C), this time considered over the field of real numbers.

Let us now turn to the reducible subalgebras. The metric is indefinite, so a vector space is characterized by its dimension and signature, i.e., the number of mutually orthogonal

TABLE III. The compact Lie algebra $g_2^C(\mathbf{R})$ and its maximal subalgebras.

Algebra	Real dimension	Basis and matrix realization	Invariant subspace (metric I_7)
$g_2^C(\mathbf{R}) \sim g_2(\mathbf{C}) \cap O(7)$	14	(3.8) and (4.14) with $M_{ik} \in \mathbf{R}$	...
$su(2) \sim sl(2,\mathbf{C}) \cap g_2^C(\mathbf{R})$	3	$i(h_{\alpha_1} + h_{\alpha_2}), \epsilon_{\alpha_1} + \epsilon_{\alpha_2}, \eta_{\alpha_1} + \eta_{\alpha_2}$ and (4.43)	...
$su(3) \sim g_2^C(\mathbf{R}) \cap o(6)$	8	(4.46) and (4.44)	$V_1^T = \{0\,0\,0\,x\,0\,0\,0\}$
$su(2) \oplus su(2) \sim g_2^C(\mathbf{R}) \cap [o(4) \oplus o(3)]$	6	(4.48) and (4.33) with $y_i, z_i \in \mathbf{R}$	$V_3^T = \{0\,0\,0\,0\,x\,y\,z\}$

basis vectors of positive $(+)$, or negative $(-)$, or zero (0) length. Let us now run through all possibilities.

We start by considering nondegenerate invariant subspaces and use the diagonal realization (4.18), with all entries real and the metric (4.17).

(1) $(+)$. We choose this vector space to be $V_+^T = (0\,0\,0\,x\,0\,0\,0)$ and obtain $m_i = n_i = 0$, i.e.,

$$M = \begin{pmatrix} (A - A^T)/2 & 0 & -(A + A^T)/2 \\ 0 & 0 & 0 \\ -(A + A^T)/2 & 0 & (A - A^T)/2 \end{pmatrix},$$

$$A_{ik} \in \mathbf{R}, \quad \operatorname{Tr} A = 0, \tag{4.50}$$

which is readily identified as $sl(3,\mathbf{R})$. It coincides with (2.15), viewed over $\mathbf{R}$. We have

$$sl(3,\mathbf{R}) \sim g_2^{NC}(\mathbf{R}) \cap O(3,3). \tag{4.51}$$

(2) $(-)$. We choose this vector space as

$$V_-^T = (0\,0\,0\,0\,0\,0\,x). \tag{4.52}$$

Requiring $MV_- \subseteq V_-$ we obtain a su(2,1) subalgebra, realized by the o(4,2) matrices:

$$M = \begin{pmatrix} 0 & 2(c_3 - c_0) & -2c_2 & 2c_1 & -n_3 & -n_4 & 0 \\ 2(c_0 - c_3) & 0 & 2c_1 & 2c_2 & -n_4 & n_3 & 0 \\ 2c_2 & -2c_1 & 0 & 2(c_0 + c_3) & -2n_2 & -2n_1 & 0 \\ -2c_1 & -2c_2 & -2(c_0 + c_3) & 0 & 2n_1 & 2n_2 & 0 \\ -n_3 & -n_4 & -2n_2 & 2n_1 & 0 & -4c_0 & 0 \\ -n_4 & n_3 & -2n_1 & 2n_2 & 4c_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{4.53}$$

(the c_i and n_i correspond to compact and noncompact generators, respectively). We have in this case

$$su(2,1) \sim g_2^{NC}(\mathbf{R}) \cap O(4,2). \tag{4.54}$$

Equivalently, we could realize

$$su(2,1) \sim g_2^{NC}(\mathbf{R}) \cap su(4,2). \tag{4.55}$$

In the canonical basis (3.10) for $g_2^{NC}(\mathbf{R})$ one realization of (4.55) is

$$su(2,1) \sim \{ih_{\alpha_1}, ih_{\alpha_2}, \epsilon_{\alpha_1}, \eta_{\alpha_1}, i\epsilon_{\alpha_1 + 3\alpha_1}, i\eta_{\alpha_1 + 3\alpha_2}, i\epsilon_{2\alpha_1 + 3\alpha_2}, i\eta_{2\alpha_1 + 3\alpha_2}\}. \tag{4.56}$$

(3) $(++)$, $(--)$, or $(+-)$. Similarly as in the case of $g_2(\mathbf{C})$ two-dimensional nondegenerate invariant subspaces do not lead to maximal subalgebras.

(4) $(---)$. We choose the space in the form

$$V_{---}^T = (0\,0\,0\,0\,x\,y\,z). \tag{4.57}$$

The invariance of V_{---}^T implies $n = 0$, $V = 0$ in (4.18) and we are left with the maximal compact subalgebra $su(2) \oplus su(2)$ of $g_2^{NC}(\mathbf{R})$. In this case we have

$$su(2) \oplus su(2) \sim g_2^{NC}(\mathbf{R}) \cap o(4) \oplus o(3). \tag{4.58}$$

The matrix realization is (4.33) (with $y_i, z_i \in \mathbf{R}$) and in terms of the canonical basis this is conjugate to (3.10b).

(5) $(++-)$. We choose the space in the form

$$V_{++-}^T = (0\,0\,x\,y\,0\,0\,z) \tag{4.59}$$

and obtain

$$su(1,1) \oplus su(1,1) \sim g_2^{NC}(\mathbf{R}) \cap o(2,2) \oplus o(2,1). \tag{4.60}$$

In the canonical basis (3.10) we can identify this algebra, e.g., as

$$su(1,1) \oplus su(1,1) \sim \{ih_{\alpha_2}, i\epsilon_{\alpha_2}, i\eta_{\alpha_2}\} \oplus \{ih_{2\alpha_1 + 3\alpha_2}, i\epsilon_{2\alpha_1 + 3\alpha_2}, i\eta_{2\alpha_1 + 3\alpha_2}\}. \tag{4.61}$$

(6) $(+--)$. We choose

$$V_{+--}^T = (0\,0\,0\,x\,y\,z\,0). \tag{4.62}$$

The invariance of this subspace leads to an o(2,1) subalgebra, contained in $sl(3,\mathbf{R})$ and hence not maximal.

(7) $(+++)$. The invariance of such a subspace leads to an o(3,1) subalgebra that is not maximal.

The remaining subspaces to be considered are degenerate, i.e. their bases contain at least one isotropic vector. In order to lead to a maximal subalgebra such a space must be completely isotropic. We shall consider these subspaces in the metric J_7 (4.19) and hence use the realization (4.21)

TABLE IV. The noncompact real Lie algebra $g_2^{\mathrm{NC}}(\mathbf{R})$ and its maximal subalgebras.

Algebra	Real dimension	Basis and matrix realization	Invariant subspace and metric used
$g_2^{\mathrm{NC}}(\mathbf{R}) \sim g_2(\mathbb{C}) \cap o(4,3)$	14	(2.4) over **R**, or (3.10);(4.10)	...
$\sim g_2(\mathbb{C}) \cap su(4,3)$		(4.18) or (4.21) over **R**	
$su(1,1) \sim sl(2,\mathbb{C}) \cap g_2^{\mathrm{NC}}(\mathbb{R})$	3	(4.49)	...
$sl(3,\mathbf{R}) \sim g_2^{\mathrm{NC}}(\mathbf{R}) \cap o(3,3)$	8	(2.15) over **R** and (4.50)	$V_+^T = (000x000)$, $K = I_{4,3}$
$su(2,1) \sim g_2^{\mathrm{NC}}(\mathbf{R}) \cap o(4.2)$	8	(4.56) and (4.53)	$V_-^T = (000000x)$, $K = I_{4,3}$
$su(2) \oplus su(2)$ $\sim g_2^{\mathrm{NC}}(\mathbf{R}) \cap [o(4) \oplus o(3)]$	6	(3.10b) and (4.33) over $\mathbb{R}$	$V_{---}^T = (0000xyz)$, $K = I_{4,3}$
$su(1,1) \oplus su(1,1)$ $\sim g_2^{\mathrm{NC}}(\mathbb{R}) \cap [o(2,2) \oplus o(2,1)]$	6	$\{ih_{\alpha_2}, i\epsilon_{\alpha_2}, i\eta_{\alpha_2}\} + \{i(2h_{\alpha_1} + 3h_{\alpha_2}), i\epsilon_{2\alpha_1+3\alpha_2}, i\eta_{2\alpha_1+3\alpha_2}\}$	$V_{++-}^T = (00xy00z)$, $K = I_{4,3}$
$P_{\alpha_1}(\mathbf{R}) \sim g_2^{\mathrm{NC}}(\mathbb{R}) \cap sim(3,2)$	9	(3.18) and (4.21) with (4.36) over **R**	$L_1^T = (000000x)$, $K = J_7$
$P_{\alpha_2}(\mathbf{R}) \sim g_2^{\mathrm{NC}}(\mathbb{R}) \cap opt(3,2)$	9	(3.19) and (4.21) with (4.39) over **R**	$L_2^T = (00000xy)$, $K = J_7$

with all entries real. The three possible invariant subspaces are (0), (00), and (000) and we choose them as in (4.35), (4.38), and (4.41), respectively. The resulting maximal parabolic subalgebras are exactly the same as in the case of $g_2(\mathbb{C})$, however viewed over the field of real numbers $\mathbb{R}$. To be more specific, we have the following.

(8) The space (0) is left invariant by $P_{\alpha_1}(\mathbb{R})$, i.e., (2.19) viewed over the $\mathbb{R}$. We have

$$P_{\alpha_1}(\mathbf{R}) \sim g_2^{\mathrm{NC}}(\mathbf{R}) \cap \mathrm{sim}(3,2), \tag{4.63}$$

where sim(3,2) is the corresponding maximal parabolic subalgebra of o(4,3). We recall that Sim(p,q) is the group of linear transformations of the Minkowski space $M(p,q)$ leaving the metric $ds^2 = dx_1^2 + \cdots + dx_p^2 - dx_{p+1}^2 - \cdots - dx_{p+q}^2$ invariant up to a constant scale factor: $(ds')^2 = e^\lambda\, ds^2, \lambda \in \mathbb{R}$ (see Ref. 47).

(9) The space (00) is left invariant by $P_{\alpha_2}(\mathbb{R})$, i.e., (2.20) over $\mathbb{R}$. We have

$$P_{\alpha_2}(\mathbf{R}) \sim g_2^{\mathrm{NC}}(\mathbb{R}) \cap \mathrm{opt}(3,2). \tag{4.64}$$

We recall that Opt(p,q) is a subgroup of the group of conformal transformations of the Minkowski space $M(p,q)$, leaving a lightlike vector space $\mathbf{x} - \mathbf{y}$, $(\mathbf{x} - \mathbf{y})^2 = 0$, invariant.[47]

(10) The space (000), as in the complex case, does not lead to a new maximal subalgebra.

All relevant information on the subalgebras of $g_2^{\mathrm{NC}}(\mathbb{R})$ is summarized in Table IV. We see that results for parabolic subalgebras are greatly simplified if the appropriate basis is chosen, namely one in which the metric tensor is J_7 of (4.19). The Iwasawa decomposition (3.11)–(3.13) was performed for the su(4,3) metric (4.27) and o(4,3) metric K_{C} (4.9). The subalgebra conjugate to $P_{\alpha_1}(\mathbb{R})$ would be obtained in that realization by requiring that the invariant isotropic vector space be spanned by $L^T = (0,1,-i,0,0,1,-i)$. The corresponding maximal solvable algebra in this case is given by (3.12) and (3.13).

This completes the classification of the maximal subalgebras of $g_2(\mathbb{C})$, $g_2^{\mathrm{C}}(\mathbb{R})$, and $g_2^{\mathrm{NC}}(\mathbb{R})$. The results are summarized in Tables II, III, and IV.

ACKNOWLEDGMENTS

We thank R. E. Beck and J. F. Cornwell for stimulating discussions and for calling some useful references to our attention. Thanks are also due to R. V. Moody, J. Patera, and A. Pianzola for discussions and for emphasizing the advantages of using the Chevalley basis in Lie algebraic considerations. One of the authors (V.H.) thanks J. F. Cornwell for his kind hospitality at St. Andrews.

Financial support from the "Accords culturels Belgique-Québec," making possible mutual visits is acknowledged. One of the author's (P.W.) research is partially sponsored by the Natural Sciences and Engineering Research Council of Canada and the "Fonds FCAR du Gouvernement du Québec."

[1]R. L. Anderson, Lett. Math. Phys. **4**, 1 (1980).
[2]R. L. Anderson, J. Harnad, and P. Winternitz, Lett. Math. Phys. **5**, 143 (1981); Physica D **4**, 164 (1982).
[3]J. Harnad, P. Winternitz, and R. L. Anderson, J. Math. Phys. **24**, 1062 (1982).
[4]S. Shnider and P. Winternitz, Lett. Math. Phys. **8**, 69 (1984); J. Math. Phys. **25**, 3155 (1984).
[5]M. del Olmo, M. A. Rodríguez, and P. Winternitz, J. Math. Phys. **27**, 14 (1986).
[6]P. Winternitz, in *Nonlinear Phenomena, Lecture Notes in Physics*, Vol. 189, edited by K. B. Wolf (Springer, Berlin, 1983), pp. 265–331.
[7]S. Lie and G. Sheffers, *Vorlesungen über continuierlichen Gruppen mit geometrischen und anderen Anwendungen* (Teubner, Leipzig, 1983) [reprinted (Chelsea, New York, 1967)].
[8]N. Jacobson, *Lie Algebras* (Dover, New York, 1962).
[9]N. Jacobson, *Exceptional Lie Algebras* (Marcel Decker, New York, 1971).
[10]R. E. Behrends, J. Dreitlein, C. Fronsdal, and B. W. Lee, Rev. Mod. Phys. **34**, 1 (1962).
[11]M. Günaydin and F. Gursey, Lett. Nuovo Cimento **6**, 401 (1973); J. Math. Phys. **14**, 1651 (1973); Phys. Rev. D **9**, 3387 (1974); M. Günaydin, Nuovo Cimento **29** A, 467 (1975).
[12]R. Casalbuoni, G. Domokos, and S. Kövesi-Domokos, Nuovo Cimento **31** A, 423 (1976).
[13]R. D. Schafer, *An Introduction to Nonassociative Algebras* (Academic, New York, 1966).
[14]P. Cvitanovic, Phys. Rev. D **14**, 1536 (1976); Nucl. Phys. B **188**, 373 (1981).
[15]G. Racah, Phys. Rev. **76**, 1352 (1949).
[16]B. Judd, *Second Quantization and Atomic Spectroscopy* (Johns Hopkins, Baltimore, 1966).
[17]D. T. Sviridov and Yu. F. Smirnov, *Teoriya opticheskykh spektrov ionov perekhodnykh metallov* (Nauka, Moscow, 1977) (*Theory of the Optical Spectra of Intermediate Metal Ions*).
[18]E. Cremmer and B. Julia, Nucl. Phys. B **159**, 141 (1971).
[19]B. Dorizzi, B. Grammaticos, R. Padjen, and V. Papageorgiou, J. Math. Phys. **25**, 2200 (1984).

[20]J. Beckers, V. Hussin, and P. Winternitz, Lett. Math. Phys. **11**, 81 (1986).
[21]G. Warner, *Harmonic Analysis on Semi-Simple Lie Groups* (Springer, Berlin, 1972).
[22]J. F. Cornwell, J. Math. Phys. **16**, 1992 (1975); **20**, 547 (1979).
[23]K. Iwasawa, Ann. Math. **50**, 507 (1949).
[24]R. P. Langlands, *Problems in the Theory of Automorphic Forms* in *Lectures in Modern Analysis and Applications III, Lecture Notes in Mathematics*, Vol. 170 (Springer, Berlin, 1970), pp. 18–86.
[25]C. Chevalley, Tohuku Math. J. (2) **7**, 14 (1955).
[26]R. W. Carter, *Simple Groups of Lie Type* (Wiley, New York, 1972).
[27]E. Cartan, Ann. Sci. Ecole Norm. Sup. **31**, 263 (1914); *Oeuvres complètes* (Gauthier-Villars, Paris, 1952).
[28]H. Freudenthal, Adv. Math. **1**, 143 (1964).
[29]J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1972).
[30]D. Rand, "PASCAL" programs for identification of Lie algebras 1," to appear in Comp. Phys. Commun.; D. Rand, P. Winternitz, and H. Zassenhaus, preprint, CRM-1351, Montréal, 1986.
[31]R. V. Moody and J. Patera, SIAM J. Alg. Discr. Meth. **5**, 359 (1984).
[32]V. G. Kac, *Infinite Dimensional Lie Algebras* (Birkhäuser, Boston, 1983).
[33]R. V. Moody, J. Algebra **10**, 211 (1968).
[34]H. Garland, J. Algebra **53**, 490 (1980); Publ. Math. IHES **52**, 181 (1980).
[35]D. Mitzman, thesis, State University of New Jersey at Rutgers, 1983.
[36]A. Barut and R. Rączka, *Theory of Group Representations and Applications* (PWN, Polish Scientific, Warsaw, 1977).
[37]F. Gantmacher, Mat. Sb. (N.S.) **5**, 101, 217 (1939).
[38]J. M. Ekins and J. F. Cornwell, Rep. Math. Phys. **7**, 167 (1976).
[39]J. Tits, Indag. Math. **28**, 223 (1966).
[40]Z. X. Wan, *Lie Algebras* (Pergamon, Oxford, 1975).
[41]J. F. Cornwell, *Group Theory in Physics* (Academic, London, 1984).
[42]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **15**, 1378, 1932 (1974).
[43]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **18**, 2259 (1977).
[44]E. B. Dynkin, Trans. Am. Math. Soc. (2) **6**, 111 (1957).
[45]J. F. Cornwell, Rep. Math. Phys. **2**, 239,289 (1971); **3**, 91 (1972); J. M. Ekins and J. F. Cornwell, Rep. Math. Phys. **5**, 17 (1974).
[46]J. A. Wolf, Mem. Am. Math. Soc. **180**, 1 (1976).
[47]J. Beckers, J. Harnad, M. Perroud, and P. Winternitz, J. Math. Phys. **19**, 2126 (1978).

Eulerian parametrization of Wigner's little groups and gauge transformations in terms of rotations in two-component spinors

D. Han
SASC Technologies, Inc., 5809 Annapolis Road, Hyattsville, Maryland 20784
Y. S. Kim
Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742
D. Son
Department of Physics, Kyungpook National University, Daegu 635, Korea

(Received 20 February 1986; accepted for publication 30 April 1986)

A set of rotations and Lorentz boosts is presented for studying the three-parameter little groups of the Poincaré group. This set constitutes a Lorentz generalization of the Euler angles for the description of classical rigid bodies. The concept of Lorentz-generalized Euler rotations is then extended to the parametrization of the E(2)-like little group and the O(2,1)-like little group for massless and imaginary-mass particles, respectively. It is shown that the E(2)-like little group for massless particles is a limiting case of the O(3)-like or O(2,1)-like little group. A detailed analysis is carried out for the two-component SL($2,c$) spinors. It is shown that the gauge degrees of freedom associated with the translationlike transformation of the E(2)-like little group can be traced to the SL($2,c$) spins that fail to align themselves to their respective momenta in the limit of large momentum and/or vanishing mass.

I. INTRODUCTION

The Euler angles constitute a convenient parametrization of the three-dimensional rotation group. The Euler kinematics consists of two rotations around the z axis with one rotation around the y axis between them. The first question we would like to address in this paper is what happens if we add a Lorentz boost along the z direction to this traditional procedure. Since the rotation around the z axis is not affected by the boost along the same axis, we are asking what is the Lorentz-generalized form of the rotation around the y axis.

Since the publication of Wigner's fundamental paper on the Poincaré group in 1939,[1] a number of mathematical techniques have been developed to deal with the three-parameter little groups that leave a given four-momentum invariant. Our second question is why we do not yet have a standard set of transformations for Wigner's little groups.

In this paper, we combine the first and second questions. One of Wigner's little groups is locally isomorphic to O(3). Furthermore, the Euler angles constitute the natural language for spinning tops in classical mechanics, while Wigner's little groups describe the internal space-time symmetries of relativistic particles, including spins. It is thus quite natural for us to look for a possible Eulerian parametrization of the three-parameter little groups.

As far as massive particles are concerned, the traditional approach to this problem is to go to the Lorentz frame in which the particle is at rest, and then perform rotations there.[1] Then, its four-momentum is not affected, but the direction of its spin becomes changed. This operation, however, is not possible for massless or imaginary-mass particles.

In order to construct a Lorentz kinematics that includes both massive and massless particles, we observe that the transformation that changes a given four-momentum can be carried out in many different ways. However, as Wigner observed in 1957, the resulting spin orientation depends on the way in which the transformation is performed and on the mass of the particle.[2] For instance, when a particle with positive helicity is rotated, the helicity remains unchanged. As far as the momentum is concerned, we can achieve the same purpose by performing a simple boost. However, this boost does not leave the helicity invariant. Furthermore, the change in the direction of spin depends on the mass.

Indeed, the difference between the rotation and boost was studied for massless photons by Kupersztych,[3] who observed that this difference amounts to a gauge transformation. In this paper, we extend the kinematics of Kupersztych to include massive and imaginary-mass particles. We shall show that this extended kinematics constitutes the abovementioned Lorentz generalization of the Euler rotations.

We then study the extended Kupersztych kinematics using the SL($2,c$) spinors. Among the four two-component SL($2,c$) spinors, two of them preserve the helicity under boosts in the zero-mass limit, as was noted by Wigner in 1957. However, the remaining two do not preserve the helicity in the same limit. We show that these helicity nonpreserving spinors are responsible for gauge degrees of freedom contained in the E(2)-like little group for photons.

In Sec. II, we work out the Kupersztych kinematics for massive particles. It is pointed out that this new kinematics is equivalent to the traditional O(3)-like kinematics in which the particle is rotated in its rest frame. We show in Sec. III that the E(2)-like little group for massless particles is the infinite-momentum/zero-mass limit of the O(3)-like little group discussed in Sec. II. In Sec. IV, we discuss the continuation of the transformation matrices for the O(3)-like little

 0022-2488/86/092228-08$02.50

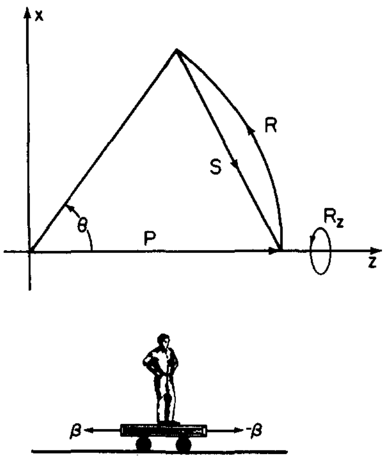


FIG. 1. Lorentz-generalized Euler rotations. The traditional Euler parametrization consists of two rotations around the z axis with one rotation around the y axis between them. If we add a Lorentz boost along the z axis, the two rotations around the z axis are not affected. The rotation around the y axis can be Lorentz-generalized in the following manner. If we boost the system along the z direction, we are dealing with the system with a nonzero four-momentum along the same direction. The four-momentum p can be rotated around the y axis by angle θ. The same result can be achieved by boost S^{-1}. However, these two transformations do not produce the same effect on the spin. The most effective way of studying this difference is to study the transformation SR, which leaves the initial four-momentum invariant.

group to the case of imaginary-mass particles.

In Sec. V, we study the transformation properties of the four two-component spinors in the SL(2,c) regime. It is shown that in the limit of infinite momentum and/or zero mass, two of the SL(2,c) spinors preserve their respective helicities, while the remaining two do not. We note, in Sec. VI, that four-vectors can be constructed from the four two-component SL(2,c) spinors. It is shown that the origin of the gauge degrees of freedom for photons can be traced to the spinors that refuse to align themselves to the momentum in the infinite-momentum/zero-mass limit.

II. KINEMATICS OF THE O(3)-LIKE LITTLE GROUP

The Euler rotation consists of a rotation around the y axis preceded *and* followed by rotations around the z axis. If the boost is made along the z axis, the rotations around the z axis are not affected. In this section, we discuss a Lorentz generalization of the rotation around the y axis and its relation to the O(3)-like little group for massive particles.

Let us start with a massive particle at rest whose four-momentum is

$$(0,0,0,m)\,. \tag{1}$$

We use the four-vector convention: $x^{\mu} = (x, y, z, t)$. We can boost the above four-momentum along the z direction with velocity parameter α:

$$P = m(0,0,\alpha/(1-\alpha^2)^{1/2}, 1/(1-\alpha^2)^{1/2})\,. \tag{2}$$

The four-by-four matrix which transforms the four-vector of Eq. (1) to that of Eq. (2) is

$$A(\alpha) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/(1-\alpha^2)^{1/2} & \alpha/(1-\alpha^2)^{1/2} \\ 0 & 0 & \alpha/(1-\alpha^2)^{1/2} & 1/(1-\alpha^2)^{1/2} \end{pmatrix}. \tag{3}$$

Let us next rotate the four-vector of Eq. (2) using the rotation matrix:

$$R(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{4}$$

This rotation does not alter the helicity of the particle.[2]

As is specified in Fig. 1, we can achieve the same result on the four-momentum by applying a boost matrix. However, unlike the rotation of Eq. (4), this boost is not a helicity-preserving transformation.[2] We can study the difference between these two transformations by taking the product of the rotation and the inverse of the boost. This inverse boost is illustrated in Fig. 1, and is represented by

$$S = \begin{pmatrix} 1 + 2(\sinh(\lambda/2)\cos(\theta/2))^2 & 0 & -(\sinh(\lambda/2))^2\sin\theta & -(\sinh\lambda)\cos(\theta/2) \\ 0 & 1 & 0 & 0 \\ -(\sinh(\lambda/2))^2\sin\theta & 0 & 1 + 2(\sinh(\lambda/2)\sin(\theta/2))^2 & (\sinh\lambda)\sin(\theta/2) \\ -(\sinh\lambda)\cos(\theta/2) & 0 & (\sinh\lambda)\sin(\theta/2) & \cosh\lambda \end{pmatrix}, \tag{5}$$

where

$$\lambda = 2[\tanh^{-1}(\alpha\sin(\theta/2))]\,. \tag{6}$$

This matrix depends on the rotation angle θ and the velocity parameter α, and becomes an identity matrix when the particle is at rest with $\alpha = 0$.

Indeed, the rotation $R(\theta)$ followed by the boost $S(\alpha,\theta)$ leaves the four-momentum p of Eq. (2) invariant:

$$P = D(\alpha,\theta)P\,, \tag{7}$$

where

$$D(\alpha,\theta) = S(\alpha,\theta)R(\theta)\,.$$

The multiplication of the two matrices is straightforward, and the result is

$$D(\alpha,\theta)=\begin{pmatrix} 1-(1-\alpha^2)u^2/2T & 0 & -u/T & \alpha u/T \\ 0 & 1 & 0 & 0 \\ u/T & 0 & 1+u^2/2T & \alpha u^2/2T \\ \alpha u/T & 0 & -\alpha u^2/2T & 1+\alpha u^2/2T \end{pmatrix}. \tag{8}$$

where

$u=-2(\tan(\theta/2))$ and $T=1+(1-\alpha^2)(\tan(\theta/2))^2$.

This complicated expression leaves the four-momentum P of Eq. (2) invariant. Indeed, if the particle is at rest with vanishing velocity parameter α, the above expression becomes a rotation matrix. As the velocity parameter α increases, this D matrix performs a combination of rotation and boost, but leaves the four-momentum invariant.

Let us approach this problem in the traditional framework.[1] The above transformation is clearly an element of the O(3)-like little group that leaves the four-momentum P invariant. Then we can boost the particle with its four-momentum P by A^{-1} until the four-momentum becomes that of Eq. (1), rotate it around the y axis, and then boost it by A until the four-momentum becomes P of Eq. (2). It is appropriate to call this rotation in the rest frame the *Wigner rotation*.[4] The transformation of the O(3)-like little group constructed in this manner should take the form

$$D(\alpha,\theta)=A(\alpha)W(\theta^*)[A(\alpha)]^{-1}, \tag{9}$$

where W is the Wigner rotation matrix

$$W(\theta^*)=\begin{pmatrix} \cos\theta^* & 0 & \sin\theta^* & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta^* & 0 & \cos\theta^* & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{10}$$

We may call θ^* the *Wigner angle*. The question then is whether D of Eq. (9) is the same as D of Eq. (8). In order to answer this question, we first take the trace of the expression given in Eq. (9). The similarity transformation of Eq. (9) assures us that the trace of W be equal to that of D. This leads to

$$\theta^*=\cos^{-1}\left(\frac{1-(1-\alpha^2)(\tan(\theta/2))^2}{1+(1-\alpha^2)(\tan(\theta/2))^2}\right). \tag{11}$$

It is then a matter of matrix algebra to confirm that D of Eq. (9) and that of Eq. (8) are identical.

We have plotted in Fig. 2 the Wigner rotation angle θ^* as a function of the velocity parameter α. Here θ^* becomes θ when $\alpha=0$, and remains approximately equal to θ when α is smaller than 0.4. Then θ^* vanishes when $\alpha\to 1$. Indeed, for a given value of θ, it is possible to determine the value of θ^* that is the rotation angle in the Lorentz frame in which the particle is at rest.

The D matrix in the traditional form of Eq. (9) is well known.[1] However, the fact that it can also be derived from the closed-loop $R(\theta)$ and $S(\alpha,\theta)$ suggests that it has a richer content. For instance, the closed-loop kinematics does not have to be unique. There is at least one other closed-loop kinematics that leaves the four-momentum invariant.[5] The Kupersztych kinematics, which we are using in this paper, is convenient for studying the relation between the Euler angles and the parameters of the O(3)-like little group.

We have so far discussed the transformations in the x-z plane. It is quite clear that the same analysis can be carried out in the y-z plane or any other plane containing the z axis. This means that we can perform rotations $R_z(\phi)$ and $R_z(\psi)$, respectively, before and after carrying out the transformations in the x-z plane. Indeed, together with the velocity parameter α, the three parameters θ, ϕ, and ψ constitute the Eulerian parametrization of the O(3)-like little group.

III. E(2)-LIKE LITTLE GROUP FOR MASSLESS PARTICLES

Let us study in this section the D matrix of Eq. (8) as the particle mass becomes vanishingly small, by taking the limit of $\alpha\to 1$. In this limit, the D matrix of Eq. (8) becomes

$$D(u)=\begin{pmatrix} 1 & 0 & -u & u \\ 0 & 1 & 0 & 0 \\ u & 0 & 1-u^2/2 & u^2/2 \\ u & 0 & -u^2/2 & 1+u^2/2 \end{pmatrix}. \tag{12}$$

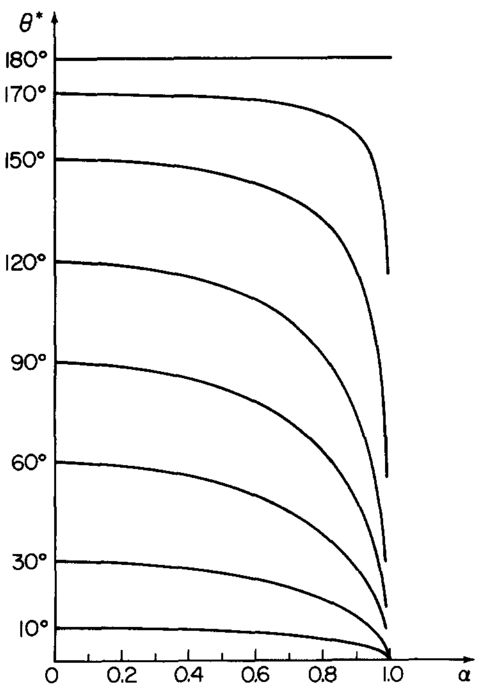


FIG. 2. Wigner rotation angle versus lab-frame rotation angle. We have plotted θ^* as a function of α for various values of θ using Eq. (11). $\theta=\theta^*$ at $\alpha=0$. θ^* is nearly equal to θ for moderate values of α, but it rapidly approaches 0 as α becomes 1.

After losing the memory of how the zero-mass limit was taken, it is impossible to transform this matrix into a rotation matrix. There is no Lorentz frame in which the particle is at rest. If we boost this expression along the z direction using the boost matrix

$$B(\beta)=\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1/(1-\beta^2)^{1/2} & \beta/(1-\beta^2)^{1/2}\\ 0 & 0 & \beta/(1-\beta^2)^{1/2} & 1/(1-\beta^2)^{1/2}\end{pmatrix}, \tag{13}$$

D remains form-invariant:

$$D'(u)=B(\beta)D(u)[B(\beta)]^{-1}=D(u'), \tag{14}$$

where

$$u'=[(1+\beta)/(1-\beta)]^{1/2}u .$$

The matrix of Eq. (12) is the case where the Kupersztych kinematics is performed in the x-z plane. This kinematics also can be performed in the y-z plane. Thus the most general form for the D matrix is

$$D(u,v)=\begin{pmatrix}1 & 0 & -u & u\\ 0 & 1 & -v & v\\ u & v & 1-(u^2+v^2)/2 & (u^2+v^2)/2\\ u & v & -(u^2+v^2)/2 & 1+(u^2+v^2)/2\end{pmatrix}. \tag{15}$$

The algebraic property of this expression has been discussed extensively in the literature.[1,5–8] If applied to the photon four-potential, this matrix performs a gauge transformation.[5,7] The reduction of the above matrix into the three-by-three matrix representing a finite-dimensional representation of the two-dimensional Euclidean group has also been discussed in the literature.[8]

Let us go back to Eq. (9). We have obtained the above gauge transformation by boosting the rotation matrix W given in Eq. (10). This means that the Lorentz-boosted rotation becomes a gauge transformation in the infinite-momentum and/or zero-mass limit. This observation was made earlier in terms of the *group contraction* of O(3) to E(2),[9,10] which is a singular transformation. We are then led to the question of how the method used in this section can be analytic, while the traditional method is singular.

The answer to this question is very simple. The group contraction is a language of Lie groups.[9,10] The parameter α we use in this paper is not a parameter of the Lie group. If we use η as the Lie-group parameter for boost along the z direction, it is related to α by $\sinh\eta=\alpha/(1-\alpha^2)^{1/2}$. However, this expression is singular at $\alpha=\pm 1$. Therefore, the continuation in α is not necessarily singular. We shall continue the discussion of this limiting process in terms of the SL$(2,c)$ spinors in Sec. VI.

IV. O(2,1)-LIKE LITTLE GROUP FOR IMAGINARY-MASS PARTICLES

We are now interested in transformations that leave the four-vector of the form

$$P=im(0,0,\alpha/(\alpha^2-1)^{1/2},1/(\alpha^2-1)^{1/2}) \tag{16}$$

invariant, with α greater than 1. Although particles with imaginary mass are not observed in the real world, the transformation group that leaves the above four-momentum invariant is locally isomorphic to O(2,1) and plays a pivotal role in studying noncompact groups and their applications in physics. This group has been discussed extensively in the literature.[11]

We are interested here in the question of whether the D matrix constructed in Secs. II and III can be analytically continued to $\alpha>1$. Indeed, we can perform the rotation and boost of Fig. 1 to obtain the D matrix of the form given in Eq. (8), if α is smaller than α_0 where

$$\alpha_0^2=[1+(\tan(\theta/2))^2]/(\tan(\theta/2))^2 . \tag{17}$$

As α increases, some elements of the D matrix become singular when T vanishes or $\alpha=\alpha_0$. Mathematically, this is a simple pole that can be avoided either clockwise or counterclockwise. However, the physics of this continuation process requires a more careful investigation.

One way to study the D transformation more effectively is to boost the spacelike four-vector of Eq. (16) along the z direction to a simpler vector

$$(0,0,im,0), \tag{18}$$

using the boost matrix of Eq. (13) with the boost parameter $\beta=1/\alpha$. Consequently, the D matrix is a Lorentz-boosted form of a simpler matrix F:

$$D=B(1/\alpha)F(\lambda)[B(1/\alpha)]^{-1} . \tag{19}$$

Here F is a boost matrix along the x direction:

$$F(\lambda)=\begin{pmatrix}\cosh\lambda & 0 & 0 & \sinh\lambda\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ \sinh\lambda & 0 & 0 & \cosh\lambda\end{pmatrix}, \tag{20}$$

where

$$\begin{aligned}\tanh\lambda&=\frac{-2(\alpha^2-1)^{1/2}\tan(\theta/2)}{1+(\alpha^2-1)(\tan(\theta/2))^2},\\ \cosh\lambda&=\frac{1+(\alpha^2-1)(\tan(\theta/2))^2}{1-(\alpha^2-1)(\tan(\theta/2))^2}.\end{aligned} \tag{21}$$

If we add the rotational degree of freedom around the z axis, the above result is perfectly consistent with Wigner's original observation that the little group for imaginary-mass particles is locally isomorphic to O(2,1).[1]

We have observed earlier that the D matrix of Eq. (8) can be analytically continued from $\alpha=1$ to $1<\alpha<\alpha_0$. At $\alpha=\alpha_0$, some of its elements are singular. If $\alpha>\alpha_0$, $\cosh\lambda$ in Eqs. (20) and (21) become negative, and this is not acceptable.

One way to deal with this problem is to take advantage of the fact that the expression for $\tanh\lambda$ in Eq. (21) is never singular for real α greater than 1. This is possible if we change the signs of both $\sinh\lambda$ and $\cosh\lambda$ when we jump from $\alpha<\alpha_0$ to $\alpha>\alpha_0$. Indeed, the continuation is possible if it is accompanied by the reflection of x and t coordinates. After taking into account the reflection of the x and t coordinates, we can construct the D matrix by boosting F of Eq. (20). The expression for the D matrix for $\alpha>\alpha_0$ becomes

$$D=\begin{pmatrix} 1-2/T & 0 & u/T & -\alpha u/T \\ 0 & 1 & 0 & 0 \\ -u/T & 0 & 1+2/[(\alpha^2-1)T] & 2\alpha/[(\alpha^2-1)T] \\ -\alpha u/T & 0 & -2\alpha/[(\alpha^2-1)T] & 1-2\alpha^2[(\alpha^2-1)T] \end{pmatrix}. \tag{22}$$

This expression cannot be used for the $\alpha\to 1$ limit, but can be used for the $\alpha\to\infty$ limit. In the limit $\alpha\to\infty$, P of Eq. (16) becomes identical to Eq. (18), and the above expression becomes an identity matrix. As for the question of whether D of Eq. (22) is an analytic continuation of Eq. (8), the answer is "no," because the transition from Eq. (22) to Eq. (8) requires the reflection of the x and t axes.

V. PARTICLES WITH SPIN-$\frac{1}{2}$

The purpose of this section is to study the D kinematics of spin-$\frac{1}{2}$ particles within the framework of SL(2,c). Let us study the Lie algebra of SL(2,c) (see Refs. 12 and 13):

$$[S_i,S_j]=i\epsilon_{ijk}S_k\,,\quad [S_i,K_j]=i\epsilon_{ijk}K_k\,,$$
$$[K_i,K_j]=-i\epsilon_{ijk}S_k\,, \tag{23}$$

where S_i and K_i are the generators of rotations and boosts, respectively. The above commutation relations are not invariant under the sign change in S_i, but they remain invariant under the sign change in K_i. For this reason, while the generators of rotations are $S_i=\frac{1}{2}\sigma_i$, the boost generators can take two different signs $K_j=(\pm)(i/2)\sigma_j$.

Let us start with a massive particle at rest, and the usual normalized Pauli spinors χ_+ and χ_- for the spin in the positive and negative z directions, respectively. If we take into account Lorentz boosts, there are four spinors. We shall use the notation $\chi_\pm$ to which the boost generators $K_i=(i/2)\sigma_i$ are applicable, and $\dot\chi_\pm$ to which $K_i=-(i/2)\sigma_i$ are applicable. There are therefore four independent SL(2,c) spinors.[12,13] In the conventional four-component Dirac equation, only two of them are independent, because the Dirac equation relates the dotted spinors to the undotted counterparts. However, the recent development in supersymmetric theories,[14] as well as some of more traditional approaches,[15] indicates that both physics and mathematics become richer in the world where all four of SL(2,c) spinors are independent. In the Appendix, we examine the nature of the restriction the Dirac equation imposes on the four SL(2,c) spinors.

As Wigner did in 1957,[2] we start with a massive particle whose spin is initially along the direction of the momentum. The boost matrix, which brings the SL(2,c) spinors from the zero-momentum state to that of p, is

$$A^{(\pm)}(\alpha)=\begin{pmatrix} ((1\pm\alpha)/(1\mp\alpha))^{1/4} & 0 \\ 0 & ((1\mp\alpha)/(1\pm\alpha))^{1/4} \end{pmatrix}, \tag{24}$$

where the superscripts (+) and (−) are applicable to the undotted and dotted spinors, respectively. In the Lorentz frame in which the particle is at rest, there is only one rotation applicable to both sets of spinors. The rotation matrix corresponding to W of Eq. (10) is

$$W(\theta^*)=\begin{pmatrix} \cos(\theta^*/2) & -\sin(\theta^*/2) \\ \sin(\theta^*/2) & \cos(\theta^*/2) \end{pmatrix}, \tag{25}$$

where the rotation angle θ^* is given in Eq. (11).

Using the formula of Eq. (9), we can calculate the D matrix for the SL(2,c) spinors. The D matrix applicable to the undotted spinors is

$$D^{(+)}(\alpha,\theta)=\begin{pmatrix} 1/\sqrt{T} & (1+\alpha)u/2\sqrt{T} \\ -(1-\alpha)u/2\sqrt{T} & 1/\sqrt{T} \end{pmatrix}, \tag{26}$$

where T and u are given in Eq. (8). The D matrix applicable to the dotted spinors is

$$D^{(-)}(\alpha,\theta)=\begin{pmatrix} 1/\sqrt{T} & (1-\alpha)u/2\sqrt{T} \\ -(1+\alpha)u/2\sqrt{T} & 1/\sqrt{T} \end{pmatrix}. \tag{27}$$

We can obtain $D^{(-)}$ from $D^{(+)}$ by changing the sign of α. Both $D^{(+)}$ and $D^{(-)}$ become W of Eq. (25) when $\alpha=0$.

If the D transformation is applied to the $\chi_\pm$ and $\dot\chi_\pm$ spinors,

$$\chi'_\pm=D^{(+)}\chi_\pm\,,\quad \dot\chi'_\pm=D^{(-)}\dot\chi_\pm\,, \tag{28}$$

the angle between the momentum and the directions of the spins represented by χ_+ and $\dot\chi_-$ is

$$\theta'=\tan^{-1}((1-\alpha)\tan(\theta/2))\,, \tag{29}$$

which becomes zero as $\alpha\to 1$. On the other hand, in the case of χ_- and $\dot\chi_+$, the angle becomes

$$\theta''=\tan^{-1}((1+\alpha)\tan(\theta/2))\,. \tag{30}$$

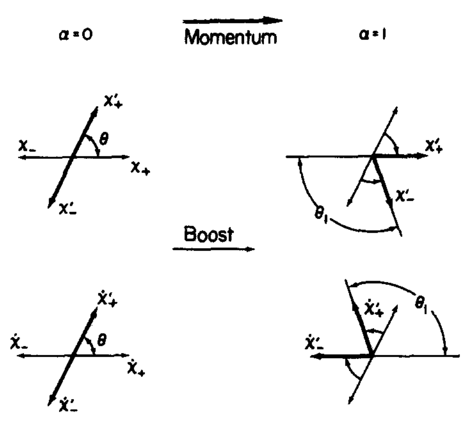


FIG. 3. Lorentz-boosted rotations of the four SL(2,c) spinors, If the particle velocity is zero, all the spinors rotate like the Pauli spinors. As the particle speed approaches that of light, two of the spins line up with the momentum, while the remaining two refuse to do so. Those spinors that line up are gauge-invariant spinors. Those that do not are not gauge invariant, and they form the origin of the gauge degrees of freedom for photon four-potentials.

In the limit of $\alpha \to 1$, this angle becomes θ_1, where

$$\theta_1 = \tan^{-1}(2(\tan(\theta/2))) . \tag{31}$$

Indeed, the spins represented by χ_- and $\dot{\chi}_+$ refuse to align themselves with the momentum. This result is illustrated in Fig. 3.

There are D transformations for the $\alpha > 1$ case. In the special Lorentz frame in which the four-momentum takes the form of Eq. (18), the D transformation becomes that of a pure boost along the x axis:

$$F^{(\pm)}(\lambda) = \begin{pmatrix} \cosh(\lambda/2) & \pm\sinh(\lambda/2) \\ \pm\sinh(\lambda/2) & \cosh(\lambda/2) \end{pmatrix}, \tag{32}$$

where λ is given in Eq. (21).

For $\alpha < \alpha_0$, we can continue to use $D^{(+)}$ and $D^{(-)}$ given in Eq. (26) and Eq. (27), respectively. However, for $\alpha > \alpha_0$, the D matrix is

$$D^{(\pm)}(\alpha,\theta) = \begin{pmatrix} (\alpha^2-1)^{1/2}(\tan(\theta/2))/\sqrt{-T} & \pm((\alpha\pm1)/(\alpha\mp1))^{1/2}/\sqrt{-T} \\ \pm((\alpha\mp1)/(\alpha\pm1))^{1/2}/\sqrt{-T} & (\alpha^2-1)^{1/2}(\tan(\theta/2))/\sqrt{-T} \end{pmatrix}. \tag{33}$$

The above expression becomes an identity matrix when $\alpha \to \infty$, as is expected from the result of Sec. IV. The D matrices of Eq. (33) are not analytic continuations of their counterparts given in Eqs. (26) and (27), because the continuation procedure, which we adopted in Sec. IV and used in this section, involves reflections in the x and t coordinates.

VI. GAUGE TRANSFORMATIONS IN TERMS OF ROTATIONS OF SPINORS

It is clear from the discussions of Secs. III–V that the limit $\alpha \to 1$ can be defined from both directions, namely from $\alpha < 1$ and from $\alpha > 1$. In the limit $\alpha \to 1$, $D^{(+)}$ and $D^{(-)}$ of Eq. (26) and Eq. (27) become

$$D^{(+)} = \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}, \quad D^{(-)} = \begin{pmatrix} 1 & 0 \\ -u & 1 \end{pmatrix}. \tag{34}$$

After going through the same procedure as that from Eq. (12) to Eq. (15), we arrive at the gauge transformation matrices[8]

$$D^{(+)}(u,v) = \begin{pmatrix} 1 & u-iv \\ 0 & 1 \end{pmatrix}, \qquad D^{(-)}(u,v) = \begin{pmatrix} 1 & 0 \\ -u-iv & 1 \end{pmatrix}, \tag{35}$$

applicable to the SL(2,c) spinors, where the $D^{(\pm)}$ are applicable to undotted and dotted spinors, respectively.

The SL(2,c) spinors are gauge invariant in the sense that

$$D^{(+)}(u,v)\chi_+ = \chi_+ , \quad D^{(-)}(u,v)\dot{\chi}_- = \dot{\chi}_- . \tag{36}$$

On the other hand, the SL(2,c) spinors are gauge dependent in the sense that

$$D^{(+)}(u,v)\chi_- = \chi_- + (u-iv)\chi_+ , \qquad D^{(-)}(u,v)\dot{\chi}_+ = \dot{\chi}_+ - (u+iv)\dot{\chi}_- . \tag{37}$$

The gauge-invariant spinors of Eq. (36) appear as polarized neutrinos in the real world. However, where do the above gauge-dependent spinors stand in the physics of spin-$\frac{1}{2}$ particles? Are they really responsible for the gauge dependence of electromagnetic four-potentials when we construct a four-vector by taking a bilinear combination of spinors?

The relation between the SL(2,c) spinors and the four-vectors has been discussed for massive particles. However, it is not yet known whether the same holds true for the massless case. The central issue is again the gauge transformation. The four-potentials are gauge dependent, while the spinors allowed in the Dirac equation are gauge invariant. Therefore, it is not possible to construct four-potentials from the Dirac spinors.

On the other hand, there are gauge-dependent SL(2,c) spinors, which are given in Eq. (37). They disappear from the Dirac spinors because N_- vanishes in the $\alpha \to 1$ limit. However, these spinors can still play an important role if they are multiplied by N_+, which neutralizes N_-. Indeed, we can construct unit vectors in the Minkowskian space by taking the direct products of two SL(2,c) spinors

$$-\chi_+\dot{\chi}_+ = (1,i,0,0) , \quad \chi_-\dot{\chi}_- = (1,-i,0,0) , \qquad \chi_+\dot{\chi}_- = (0,0,1,1) , \quad \chi_-\dot{\chi}_+ = (0,0,1,-1) . \tag{38}$$

These unit vectors in one Lorentz frame are not the unit vectors in other frames. For instance, if we boost a massive particle initially at rest along the z direction, $|\chi_+\dot{\chi}_+\rangle$ and $|\chi_-\dot{\chi}_-\rangle$ remain invariant. However, $|\chi_+\dot{\chi}_-\rangle$ and $|\chi_-\dot{\chi}_+\rangle$ acquire the constant factors $[(1+\alpha)/(1-\alpha)]^{1/2}$ and $[(1-\alpha)/(1+\alpha)]^{1/2}$, respectively. We can therefore drop $|\chi_-\dot{\chi}_+\rangle$ when we go through the renormalization process of replacing the coefficient $[(1+\alpha)/(1-\alpha)]^{1/2}$ by 1 for particles moving with the speed of light.

The $D(u,v)$ matrix for the above spinor combinations should take the form

$$D(u,v) = D^{(+)}(u,v)D^{(-)}(u,v) , \tag{39}$$

where $D^{(+)}$ and $D^{(-)}$ are applicable to the first and second spinors of Eq. (38), respectively. Then

$$D(u,v)(-|\chi_+\dot{\chi}_+\rangle) = |\chi_+\dot{\chi}_+\rangle + (u+iv)|\chi_+\dot{\chi}_-\rangle , \qquad D(u,v)|\chi_-\dot{\chi}_-\rangle = |\chi_-\dot{\chi}_-\rangle + (u-iv)|\chi_+\dot{\chi}_-\rangle , \qquad D(u,v)|\chi_+\dot{\chi}_-\rangle = |\chi_+\dot{\chi}_-\rangle . \tag{40}$$

The first two equations of the above expression correspond to the gauge transformations on the photon polarization vectors. The third equation describes the effect of the D transformation on the four-momentum, confirming the fact that $D(u,v)$ is an element of the little group. The above operation is identical to that of the four-by-four D matrix of Eq. (15) on photon polarization vectors.

VII. CONCLUDING REMARKS

We studied in this paper Wigner's little groups by constructing a Lorentz kinematics that leaves the four-momen-

	Massive Slow	between	Massless Fast
Energy Momentum	$E=\frac{p^2}{2m}$	Einstein's $E=\sqrt{m^2+p^2}$	$E=p$
Spin, Gauge Helicity	S_3 S_1 S_2	Wigner's Little Group	S_3 Gauge Trans.

FIG. 4. Significance of the concept of Wigner's little groups. The beauty of Einstein's special relativity is that the energy-momentum relation for massive and slow particles and that for massless particles can be unified. Wigner's concept of the little groups unifies the internal space-time symmetries of massive and massless particles.

tum of a particle invariant. This kinematics consists of one rotation followed by one boost. Although the net transformation leaves the four-momentum invariant, the particle spin does not remain unchanged. The departure from the original spin orientation is studied in detail.

For a massive particle, this departure can be interpreted as a rotation in the Lorentz frame in which the particle is at rest. For massless particles with spin-1, the net result is a gauge transformation. For a spin-$\frac{1}{2}$ particle, there are four independent spinors as the Dirac equation indicates. As the particle mass approaches zero, the spin orientations of two of the spinors remain invariant. However, the remaining two spinors do not. It is shown that this noninvariance is the cause of the gauge degrees of freedom massless particles with spin-1.

In 1957,[2] Wigner considered the possibility of unifying the internal space-time symmetries of massive and massless particles by noting the difference between rotations and boosts. Wigner considered the scheme of obtaining the internal symmetry by taking the massless limit of the internal space-time symmetry groups for massive particles. In the present paper, we have added the gauge degrees of freedom and spinors that refuse to align themselves to the momentum in the massless limit. The result of the present paper can be summarized in Fig. 4. While Einstein's special relativity unifies the energy-momentum relations for massive and massless particles, Wigner's little group unifies the internal space-time symmetries of massive and massless particles.

ACKNOWLEDGMENTS

We are grateful to Professor Eugene P. Wigner for a very illuminating discussion on his 1957 paper[2] on transformations that preserve helicity and those that do not. We would like to thank Dr. Avi I. Hauser for explaining to us the content of his paper on possible imaginary-mass neutrinos.[14]

APPENDIX: SL(2,*c*) SPINORS IN THE DIRAC SPINORS

We pointed out in Sec. V that the four-component Dirac equation puts a restriction on the SL(2,c) spinors. Let us see how this restriction manifests itself in the limit procedure of $\alpha \to 1$. In the Weyl representation of the Dirac equation, the rotation and boost generators take the form

$$S_i = \begin{pmatrix} \frac{1}{2}\sigma_i & 0 \\ 0 & \frac{1}{2}\sigma_i \end{pmatrix}, \quad K_i = \begin{pmatrix} (i/2)\sigma_i & 0 \\ 0 & -(i/2)\sigma_i \end{pmatrix}. \tag{A1}$$

These generators accommodate both signs of the boost generators for the SL(2,c) spinors. In this representation, γ_5 is diagonal, and its eigenvalue determines the sign of the boost generators.

In the Weyl representation, the D matrix should take the form

$$D(u,v) = \begin{pmatrix} D^{(+)}(u,v) & 0 \\ 0 & D^{(-)}(u,v) \end{pmatrix}, \tag{A2}$$

applicable to the Dirac spinors, which, for the particle moving along the z direction with four-momentum p, are

$$U(\mathbf{p}) = \begin{pmatrix} N_+\chi_+ \\ \pm N_-\chi_+ \end{pmatrix}, \quad V(\mathbf{p}) = \begin{pmatrix} \pm N_-\chi_- \\ N_+\chi_- \end{pmatrix}, \tag{A3}$$

where the $+$ and $-$ signs in the above expression specify positive and negative energy states, respectively. Here N_+ and N_- are the normalization constants, and

$$N_\pm = ((1 \pm \alpha)/(1 \mp \alpha))^{1/4}. \tag{A4}$$

As the momentum/mass becomes very large, N_-/N_+ becomes very small. From Eqs. (36) and (37), we can see that the large components are gauge invariant while the small components are gauge dependent. The gauge-dependent component of the Dirac spinor disappears in the $\alpha \to 1$ limit; the Dirac equation becomes a pair of the Weyl equations. If we renormalize the Dirac spinors of Eq. (A3) by dividing them by N_+, they become

$$U(\mathbf{p}) = \begin{pmatrix} \chi_+ \\ 0 \end{pmatrix}, \quad V(\mathbf{p}) = \begin{pmatrix} 0 \\ \chi_- \end{pmatrix}, \tag{A5}$$

For $\gamma_5 = \pm 1$, respectively. The gauge-dependent spinors disappear in the large-momentum/zero-mass limit. This is precisely why we do not talk about gauge transformations on neutrinos in the two-component neutrino theory.

The important point is that we can obtain the above decoupled form of spinors immediately from the most general form of spinors by imposing the gauge invariance. This means that the requirement of gauge invariance is equivalent to $\gamma_5 = 1$, as was suspected in Ref. 8.

[1]E. P. Wigner, Ann. Math. **40**, 149 (1939); V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. (U.S.A.) **34**, 211 (1946); E. P. Wigner, in *Theoretical Physics*, edited by A. Salam (I.A.E.A., Vienna, 1963).

[2]E. P. Wigner, Rev. Mod. Phys. **29**, 255 (1957). See also C. Kuang-Chao and L. G. Zastavenco, Zh. Exp. Teor. Fiz. **35**, 1417 (1958) [Sov. Phys. JETP **8**, 990 (1959)]; M. Jacob and G. C. Wick, Ann. Phys. (NY) **7**, 404 (1959).

[3]J. Kupersztych, Nuovo Cimento B **31**, 1 (1976); Phys. Rev. D **17**, 629 (1978).

[4]The Wigner rotation is frequently mentioned in the literature because two successive boosts result in a boost preceded or followed by a rotation. We believe, however, that the Wigner rotation should be defined in the Lorentz frame in which the particle is at rest, in view of the fact that the O(3)-like little group is the rotation group in the rest frame. See R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications in Physics* (Wiley, New York, 1974); A. Le Yaouanc, L. Oliver, O. Pene, and J. C. Raynal, Phys. Rev. D **12**, 2137 (1975); A. Ben-Menahem, Am. J. Phys. **53**, 62 (1985). The concept of rotations in the rest frame played an important

role in the development of quantum mechanics and atomic spectra. See L. H.Thomas, Nature **117**, 514 (1926); Philos. Mag. **3**, 1 (1927).
[5]D. Han and Y. S. Kim, Am. J. Phys. **49**, 348 (1981); D. Han, Y. S. Kim, and D. Son, Phys. Rev. D **31**, 328 (1985).
[6]E. P. Wigner, Z. Phys. **124**, 665 (1948); A. S. Wightman, in *Dispersion Relations and Elementary Particles*, edited by C. De Witt and R. Omnes (Hermann, Paris, 1960); M. Hamermesh, *Group Theory* (Addison–Wesley, Reading, MA, 1962); E. P. Wigner, in *Theoretical Physics*, edited by A. Salam (I.A.E.A., Vienna, 1962); A. Janner and T. Jenssen, Physica **53**, 1 (1971); **60**, 292 (1972); J. L. Richard, Nuovo Cimento A **8**, 485 (1972); H. P. W. Gottlieb, Proc. R. Soc. London Ser. A **368**, 429 (1979).
[7]S. Weinberg, Phys. Rev. **134**, B 882 (1964); **135**, B1049 (1964).
[8]D. Han, Y. S. Kim, and D. Son, Phys. Rev. D **26**, 3717 (1982).
[9]E. Inonu and E. P. Wigner, Proc. Natl. Acad. Sci. (U.S.A.) **39**, 510 (1953); D. W. Robinson, Helv. Phys. Acta **35**, 98 (1962); D. Korff, J. Math. Phys. **5**, 869 (1964); S. Weinberg, in *Lectures on Particles and Field Theory, Brandeis 1964*, Vol. 2, edited by S. Deser and K. W. Ford (Prentice–Hall, Englewood Cliffs, NJ, 1965); J. D. Talman, *Special Functions, A Group Theoretical Approach Based on Lectures by E. P. Wigner* (Benjamin, New York, 1968); S. P. Misra and J. Maharana, Phys. Rev. D **14**, 133 (1976).
[10]D. Han, Y. S. Kim, and D. Son, Phys. Lett. B **131**, 327 (1983); D. Han, Y. S. Kim, M. E. Noz, and D. Son, Am. J. Phys. **52**, 1037 (1984).
[11]V. Bargmann, Ann. Math. **48**, 568 (1947); L. Pukanszky, Trans. Am. Math. Soc. **100**, 116 (1961); L. Serterio and M. Toller, Nuovo Cimento **33**, 413 (1964); A. O. Barut and C. Fronsdal, Proc. R. Soc. London Ser. A **287**, 532 (1965); M. Toller, Nuovo Cimento **37**, 631 (1968); W. J. Holman and L. C. Biedenharn, Ann. Phys. (NY) **39**, 1 (1966); **47**, 205 (1968); N. Makunda, J. Math. Phys. **9**, 50, 417 (1968); **10**, 2068, 2092 (1973); K. B. Wolf, J. Math. Phys. **15**, 1295, 2102 (1974); S. Lang, SL(2,*r*) (Addison–Wesley, Reading, MA, 1975).
[12]M. A. Naimark, Am. Math. Soc. Transl. **6**, 379 (1957); I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and their Applications* (MacMillan, New York, 1963).
[13]Yu. V. Novozhilov, *Introduction to Elementary Particle Theory* (Pergamon, Oxford, 1975).
[14]S. J. Gates, M. T. Grisaru, M. Rocek, and W. Siegel, *Superspaces* (Benjamin/Cummings, Reading, MA, 1983). See also A. Chodos, A. I. Hauser, and V. A. Kostelecky, Phys. Lett. B **150**, 431 (1985); H. van Dam, Y. J. Ng, and L. C. Biedenharn, *ibid.* **158**, 227 (1985).
[15]L. C. Biedenharn, M. Y. Han, and H. van Dam, Phys. Rev. D **6**, 500 (1972).

Generating relations for reducing matrices. II. Corepresentations

M. I. Aroyo, J. N. Kotzev, and M. N. Angelova-Tjurkedjieva
Faculty of Physics, University of Sofia, Sofia, BG-1126, Bulgaria

R. Dirl and P. Kasperkovitz
Institut für Theoretische Physik, TU Wien, A-1040 Wien, Karlsplatz 13, Austria

(Received 30 October 1985; accepted for publication 9 April 1986)

The auxiliary group approach developed in Paper 1 [R. Dirl, P. Kasperkovitz, M. I. Aroyo, J. N. Kotzev, and M. Angelova-Tjurkedjieva, J. Math. Phys. 27, 37 (1986)] is generalized for the case of corepresentations of antiunitary groups. It allows us to reduce the multiplicity problem and to derive consistent generating relations for the elements of the reducing matrices for coreps. Two examples are worked out to illustrate the general scheme.

I. INTRODUCTION

In our recent paper[1] hereafter referred to as I (see also Refs. 2 and 3), we discussed the group-theoretical aspects of one of the standard problems in the theory of representations (reps) and its applications in physics: the determination of a reducing matrix S (see Refs. 4–7), which transforms a reducible unitary matrix rep R of a group G into a direct sum of irreducible reps (irreps) Γ^k,

$$S^\dagger R(g) S = \bigoplus_k \{e_k^R \times \Gamma^k(g)\}, \quad g \in G .$$

Here e_k^R is an identity matrix whose dimension equals the multiplicity of Γ^k in R, i.e., $\dim e_k^R = m_k$. According to its definition the columns of S are labeled by the triple index (k,m,a), where k is the label of the irrep Γ^k, m $(=1,\ldots,m_k)$ is the multiplicity index, and $a = 1,\ldots,n_k = \dim \Gamma^k$ is the row index of Γ^k. The square reducing matrix S can be split into rectangular submatrices $S^{k,m}$ consisting of n_R $(=\dim R)$ rows and n_k columns that satisfy

$$R(g) S^{k,m} = S^{k,m} \Gamma^k(g) , \quad g \in G .$$

For fixed Γ^k the set of blocks $S^{k,m}$, $m = 1,\ldots,m_k$, may be considered as the basis of a linear space of dimension m_k. This space becomes a unitary one if a scalar product is defined. To arrive at a unitary matrix S we assume

$$\langle S^{k,m}, S^{k',m'} \rangle = \operatorname{trace}(S^{k,m\dagger} S^{k',m'}) = n_k \delta_{m,m'} .$$

One basic problem in calculating reducing matrices is related to their nonuniqueness. This comes from the fact that every matrix S is unique only up to (i) left multiplication by unitary matrices belonging to the commuting algebra of the reducible representation R, and (ii) right multiplication by unitary matrices M belonging to the commuting algebra of the reduced representation $\oplus_k (e_k^R \times \Gamma^k)$. Because of Schur's lemma these matrices have to be of the form

$$M = \bigoplus_k M^k = \bigoplus_k (L^k \times I^k) .$$

Here I^k is an identity matrix with $\dim I^k = \dim \Gamma^k$ and L^k is an arbitrary unitary matrix whose matrix elements are labeled by the multiplicity index, $m = 1,\ldots,m_k$. The matrix L^k belongs to the commuting algebra of the rep $(e_k^R \times \Gamma^k)$.

Usually this arbitrariness inherent to the determination of S is utilized to construct a reducing matrix whose elements satisfy certain symmetry relations (e.g., symmetry under complex conjugation, permutations, associations of reps by Kronecker multiplication of a given Γ^k with one-dimensional irreps Γ^j of G, etc.). However, in most of the existing approaches to the multiplicity problem these additional symmetry requirements on the reducing matrix are applied separately and independently. Consequently the multiplicity problem remains partially unsolved in many cases where the combination of all operators would lead to a complete solution.

To gain a maximum of symmetry and generating relations in a systematic way we introduce an auxiliary group Q^{REP}. This group consists of bijective mappings of the set of all unitary matrix representations of G onto itself. Three different types of mapping are considered: (i) associations of representations, i.e., multiplication of representations with one-dimensional ones; (ii) automorphisms of representations, i.e., mappings of representations induced by automorphisms of the group G; and (iii) complex conjugation of representations. These operations are combined to form the auxiliary group Q^{REP}. For a given rep R we find a subgroup Q of Q^{REP} that leaves R invariant up to a similarity transformation and a subgroup Q^k of Q leaving Γ^k (which occurs in the reduced R) invariant up to unitary equivalence. Furthermore we define operator groups $\widetilde{Q}^k$ that are associated with Q^k and act only on the multiplicity index m of $S^{k,m}$:

$$T(q) S^{k,m} = \sum_{m'=1}^{m_k} \Delta_{m'm}(q) S^{k,m'}, \quad q \in Q^k, \quad T(q) \in \widetilde{Q}^k .$$

The space spanned by the blocks $S^{k,m}$, $m = 1,\ldots,m_k$, turns out to be a carrier space for a corepresentation (corep) Δ of the auxiliary group $\widetilde{Q}^k$ if at least one of the operations of Q^k contains the complex conjugation. In all such cases $\widetilde{Q}^k$ contains antiunitary operators and the subspaces invariant under $\widetilde{Q}^k$ are carrier spaces for coreps that are in general reducible. (The definition and the basic properties of coreps are discussed in Sec. II A. For further details see Refs. 4 and 7–10.) Therefore the resolution of the multiplicity problem is related to the reduction of the corep Δ into irreducible constituents: if Δ decomposes into inequivalent irreducible coreps the multiplicity problem is resolved, but if Δ contains an irreducible corep at least twice the multiplicity problem is solved only in part.

 0022-2488/86/092236-11$02.50

The next step of our approach consists of defining "partner blocks" $S^{k',m}$ associated to the blocks $S^{k,m}$ by generating relations of the form

$$S^{k',m} = U(q')S^{k,m} .$$

Here we assumed that $\Gamma^{k'} = q'\Gamma^k \neq \Gamma^k$, where the operation q' is a coset representative of Q with respect to Q^k and $U(q')$ the corresponding similarity transformation of R.

Like other approaches our procedure is based on the transfer of the transformation properties of the involved reps on the reducing matrices. However, our scheme is more systematic, as the various operations are closed into a group, and their combined application is often more effective than considering them separately. Our approach helps (i) to reduce (sometimes even to solve completely) the problems related to the multiplicity of Γ^k in R; and (ii) to determine generating relations for the sub-blocks $S^{k,m}$ of the reducing matrix S, which are consistent in that the effect of any sequence of the above-mentioned operations is defined and calculable. The details, the corresponding references, and three examples are given in I.

The determination of reducing matrices is the main mathematical problem in many physical applications of group theory and in particular in the well-known Wigner–Racah algebra or the method of the irreducible tensor sets (see, e.g., Refs. 4–6). This powerful technique of modern quantum mechanics is based on the theory of linear reps of groups of unitary operators. However, already in the 1930's Wigner had shown that in the physics of systems with magnetic symmetry the transformations containing the antiunitary time reversal operator play an important part. The corresponding antiunitary groups and their corepresentations (coreps) determine the transformation properties of the wave functions and operators. These specific properties also allow us to predict, for instance, degeneracies of energy levels and selection rules that can differ from those following from ordinary representation theory.

It is well known that the main theorems of the rep theory are essentially changed in the construction of the corep theory.[7–10] In the last decade this has led to intensive work on the development of the Wigner–Racah algebra for systems with magnetic symmetry on the base of the theory of the Wigner coreps.[10–17]

In this paper it is shown that the auxiliary group approach initially introduced for linear reps in I can readily be generalized to coreps. The peculiar properties of the antiunitary groups lead to a number of new symmetry properties for the elements of the reducing matrix S.

The paper is organized as follows: In Sec. II we give the scheme of the generalized auxiliary group approach for the corep case. The generating relations and uniqueness properties of the reducing matrix for coreps are discussed in Sec. III. Two examples of the application of the method for antiunitary groups for coreps are given in Sec. IV.

II. AUXILIARY GROUP APPROACH TO COREPS

A. Preliminary

For a better understanding of the peculiarities of the auxiliary group approach for coreps it is useful to start the discussion with a brief review of the main principles of the corep theory (a more detailed presentation can be found in Refs. 4 and 7–10).

Consider a Hilbert space and a group $G(H)$ of unitary/antiunitary operators acting on its elements. Here H is the subgroup of unitary operators of index 2. If a_0 is an arbitrary but fixed antiunitary element of $G(H)$, then one has the coset decomposition

$$G(H) = H + Ha_0 . \tag{2.1}$$

If we choose an orthonormal bases of the Hilbert space then the action of an operator $g\in G(H)$ onto this basis is given by an unitary matrix $D(g)$. Because of the multiplication law of the unitary/antiunitary elements of $G(H)$ the corresponding matrices have to satisfy the following composition law ("comultiplication"):

$$D(g_1)D(g_2)^{(g_1)} = D(g_1g_2) , \tag{2.2}$$

where

$$D(g_2)^{(g_1)} = \begin{cases} D(g_2) , & \text{iff } g_1 \text{ unitary,} \\ D(g_2)^* , & \text{iff } g_1 \text{ antiunitary .}\end{cases} \tag{2.3}$$

(The definition (2.3) is identical to

$$qM = \begin{cases} M , & \text{iff } q \text{ unitary ,} \\ M^* , & \text{iff } q \text{ antiunitary}\end{cases}$$

[Eq. (2.21) in I]. Nevertheless we will use both of them in order to distinguish better the origin of the complex conjugation in the expression of the type (1.3) or (2.21) of I.) According to Wigner[4] the set

$$D = \{D(g)\,|\,g\in G(H)\} , \tag{2.4}$$

endowed with a multiplication law (2.2), is called a corepresentation of $G(H)$, or "corep" for short.[4,7]

If we change the basis by a unitary transformation V, we get an equivalent corep D', which is related to D by

$$D(g) = V^\dagger D'(g) V^{(g)}, \quad g\in G(H) . \tag{2.5}$$

If the corep D is irreducible it will be denoted by D^k in the following. Each irreducible corep (coirrep) D^k of $G(H)$ is uniquely determined by the irreps Γ^k of the unitary subgroup H. According to the restriction of $D^k\downarrow H$ three types of coirreps have to be distinguished:

$$\begin{aligned} &\text{type I:} \quad D^k\downarrow H \sim \Gamma^k , \\ &\text{type II:} \quad D^k\downarrow H \sim \Gamma^k \oplus \Gamma^k , \\ &\text{type III:} \quad D^k\downarrow H \sim \Gamma^k \oplus \Gamma^{\bar{k}} , \\ &\qquad\qquad \text{where } \Gamma^{\bar{k}}(g) = \Gamma^k(a_0^{-1}ga_0)^* . \end{aligned} \tag{2.6}$$

These three types of coirreps are analogous to real, symplectic, and complex reps, respectively.

Of particular importance is also the generalization of the Schur lemma.[7,14,16] Every matrix M^k, which commutes with all matrices $D^k(g)$, $g\in G(H)$, of a coirrep D^k (in accordance with the comultiplication rule),

$$M^k D^k(g) = D^k(g) M^{k(g)} , \quad g\in G(H) , \tag{2.7}$$

is in general not a multiple of the identity matrix. The corresponding commutator algebra is a division algebra over $\mathbb{R}$ and it is isomorphic to the real numbers $\mathbb{R}$, the quaternions $\mathbb{Q}$, and the complex numbers $\mathbb{C}$ for the coreps of type I, II, and III, respectively.[12,14,16,17] This peculiarity of the generalized Schur lemma is central for the application of the corep theory.

This brief report on the corep theory will be sufficient for the generalization of the auxiliary group approach.

Following closely the general scheme developed in I, we will now discuss its modification for the corep case of antiunitary groups.

B. The auxiliary group $Q^{\rm CO}$ for coreps

The set of all bijective mappings of the coreps D of $G(H)$ onto itself, generated by the operations "association," "automorphism," and "complex conjugation," form the auxiliary group $Q^{\rm CO}$ with the structure[1]

$$Q^{\rm CO} = \mathrm{ASS} \circledS (\mathrm{AUT} \otimes \mathrm{CON}), \tag{2.8}$$

where $\circledS$ means semidirect product.

Similar to the case of reps the association of a given corep D with the Kronecker product $D^j \times D$ is defined by[1,15]

$$(a_j D)(g) = D^j(g) \times D(g), \quad g \in G(H), \quad \dim D^j = 1. \tag{2.9}$$

The corresponding operators form an Abelian subgroup ASS of $Q^{\rm CO}$.

The corep generalization of the subgroup CON, generated by the operator c of complex conjugation $\mathrm{CON} = \langle c \rangle \simeq C_2$, is also trivial,

$$(c_1 D)(g) = D(g)^{c_1} = \begin{cases} D(g)^*, & c_1 = c, \\ D(g), & c_1 = c^2 = e, \end{cases} \quad g \in G(H). \tag{2.10}$$

The differences between $Q^{\rm REP}$ and $Q^{\rm CO}$ are mainly manifested in the determination of the subgroup $\mathrm{AUT} \subset Q^{\rm CO}$, consisting of bijective mappings of the coreps of $G(H)$, generated by its automorphisms β: $G(H) \to G(H)$:

$$(b_k D)(g) = D(\beta_k^{-1}(g)), \quad g \in G(H). \tag{2.11}$$

There are two specific features that should be emphasized. First, the automorphism group Aut G of a given abstract group G is the set of all bijective mappings of G onto G that preserve the multiplication law. For the case of an antiunitary group $G(H)$ we define its automorphism group Aut $G(H)$ as the set of mappings that preserve the multiplication law *and* leave the unitary subgroup H invariant. That is, if $G \simeq G(H)$, then

$$\mathrm{Aut}\, G(H) = \{\beta \mid \beta \in \mathrm{Aut}\, G,\ G \simeq G(H),\ \beta(H) = H\} \subseteq \mathrm{Aut}\, G. \tag{2.12}$$

Second, for ordinary reps all inner automorphisms of G do not permute the classes of conjugated elements and therefore are considered as trivial transformations in $\mathrm{Aut} \subset Q^{\rm REP}$ [see the comments before Eq. (2.2) in I]. For coreps the same statement is only valid for the inner automorphisms of the unitary subgroup $H \lhd G(H)$. Hence in AUT it is necessary to include also those inner automorphism of $G(H)$ that are outer in respect to $H \lhd G(H)$. They can lead to additional symmetry relations and give more information about the structure of S. The action of the general element $Q = (a_j, b_k, c_1) \in Q^{\rm CO}$ on the coreps is determined as in Eq. (2.5) of I:

$$(qD)(g) = (a_j(b_k(c_1 D)))(g) = D^j(g) \times D(\beta_k^{-1}(g))^{c_1}. \tag{2.13}$$

This is in accordance with the semidirect multiplication rules

$$q'q = (a'; b'c')(a; bc) = (a' a^{b'c'}; b'bc'c), \tag{2.14}$$

where

$$a^{b'c'} = (b'c')a(b'c')^{-1} = b'(c'ac'^{-1})b'^{-1}. \tag{2.15}$$

Obviously if $ca = ac$ [which is the case in all gray groups and black and white groups with $a_0 \in Z(G(H))$] and if $ba = ab$ [e.g., if there is only one nontrivial one-dimensional coirrep in $G(H)$], then the automorphism (2.15) is trivial and the semidirect product in (2.14) and (2.8) becomes a direct product (see Sec. III B).

C. The groups Q and Q^k for coreps

Let $R(g)$, $g \in G(H)$, be a reducible corep that is decomposed into irreducible components D^k by a unitary transformation of the type (1.5),

$$S^\dagger R(g) S^{(g)} = \bigoplus_k (e_k^R \times D^k(g)), \quad g \in G(H). \tag{2.16}$$

Here the coirrep D^k appears $m_k = (R/D^k)$ times in identical form, i.e., e_k^R is an identity matrix with $\dim e_k^R = m_k$.

The groups Q, Q^k and the corresponding Q-equivalent classes $[k]$ are determined as in I, just by substituting "rep" with "corep" and $Q^{\rm REP}$ with $Q^{\rm CO}$:

$$Q = \{q \mid qR \sim R,\ q \in Q^{\rm CO}\} \subseteq Q^{\rm CO}, \tag{2.17}$$

$$Q^k = \{q \mid qD^k \sim D^k,\ q \in Q \subseteq Q^{\rm CO}\} \subseteq Q, \tag{2.18}$$

$$[k] = \{qD^k \mid q \in Q\} = \{q_1^{(k)} D^k, q_2^{(k)} D^k, \ldots\}. \tag{2.19}$$

Here the elements $q_1^{(k)}$ are suitable coset representatives of Q^k in Q. For convenience we adopt the following convention:

$$D^1 = q_1^{(k)} D^k, \quad \text{for } D^1 \in [k]. \tag{2.20}$$

D. The groups $\bar{Q}$ and $\bar{Q}^k$ for coreps

If $q \in Q$, there exists a set of unitary matrices $U(q)$, which relate the equivalent coreps qR and R in accordance with Eq. (2.5),

$$(qR)(g) = U(q)^\dagger R(g) U(q)^{(g)}, \quad q \in Q. \tag{2.21}$$

The appearance of $U(q)^{(g)} = U(q)^*$ for $g =$ antiunitary, does not change anything in the rest of Sec. II D in I. In particular, applying the comultiplication rule, Eq. (2.20) of I, we determine the group $\bar{Q}$, generated by the matrices $U(q)$ for a reducible corep R:

$$\bar{Q} = \langle U(q) \rangle, \quad q \in Q. \tag{2.22}$$

Analogously we define, for $q \in Q^k$,

$$(qD^k)(g) = U^k(q)^\dagger D^k(g) U^k(q)^{(g)}, \quad q \in Q^k, \tag{2.23}$$

$$\bar{Q}^k = \langle U^k(q) \rangle, \quad q \in Q^k. \tag{2.24}$$

III. THE OPERATOR GROUP $\tilde{Q}^k$ AND THE TRANSFORMATION PROPERTIES OF THE REDUCING MATRIX

Following the procedure developed in I we now define the operator group $\tilde{Q}^k$ and investigate the transformation properties of the reducing matrix S. For coreps there are some essential peculiarities, which should be discussed in detail.

The reducing matrix S was defined as a unitary matrix, which reduced a corep R into a block-diagonal form (2.16). For our further discussion it will be more convenient to split the reducing matrix S into rectangular blocks S^k consisting of all the columns of S, which belong to the coirrep D^k,

$$(S^k)^\dagger R(g) S^{k(g)} = (e_k^R \times D^k(g)), \quad g \in G(H). \tag{3.1}$$

Acting with the operators $q \in Q^k \subset Q$ on both sides of (3.1), we get

$$(qS^k)^\dagger (qR)(g)(qS^k)^{(g)} = (e_k^R \times (qD^k)(g)), \quad g \in G(H), \quad q \in Q^k. \tag{3.2}$$

From the equivalence of the coreps $qR \sim R$ [(2.21)] and $qD^k \sim D^k$ [(2.23)] follows the relation

$$(qS^k)^\dagger U(g)^\dagger R(g) U(q)^{(g)} (qS^k)^{(g)} = (e_k^R \times U^k(g)^\dagger D^k(g) U^k(q)^{(g)}). \tag{3.3}$$

Using (3.1) and the properties of the direct product of matrices, Eq. (3.3) can be transformed into the form

$$\{(S^k)^\dagger U(q)(qS^k)(e_k^R \times U^k(q))^\dagger\}(e_k^R \times D^k(g)) = (e_k^R \times D^k(g))\{(S^k)^\dagger U(q)(qS^k)(e_k^R \times U^k(q)^\dagger)\}^{(g)}, \tag{3.4}$$

where the reducible coreps of both sides of (3.4) are identical (not only equivalent).

The matrix in the curly brackets in Eq. (3.4),

$$M^k = M^k(q) = (S^k)^+ U(q)(qS^k) \times (e_k^R U^k(q))^+, \quad q \in Q^k, \tag{3.5}$$

is a unitary matrix with $\dim M^k = m_k \dim D^k$, and it commutes with the reducible corep $(e_k^R \times D^k)$. Due to the generalized Schur lemma for reducible coreps (see, for example, Ref. 14) the matrices M^k form a commutator algebra of the corep $(e_k^R \times D^k)$. The structure of this algebra is uniquely determined by the Wigner type of the coirrep D^k.

As is shown in Ref. 14, the matrices M^k can be factorized into a direct product of two submatrices,

$$M^k(q) = L^k(q) \times I^k, \quad q \in Q^k, \tag{3.6}$$

if the coirreps are in the Wigner canonical form. Here I^k is an identity matrix with $\dim I^k = \dim \Gamma^k$, where Γ^k is the corresponding unitary subgroup irrep (2.6). The matrix $L^k_{(q)}$ is an unitary matrix whose dimension equals $p_k = (R \mid \Gamma^k)$, the multiplicity of Γ^k in the restriction $R \downarrow H$. Obviously the matrices $L^k(q)$ belong to an algebra $a(L^k)$, which is isomorphic to the commutator algebra $a(M^k)$ of the reducible corep $(e_k^R \times D^k)$, i.e., $a(L^k) \simeq a(M^k)$. We shall identify $a(L^k)$ with $a(M^k)$.

For coirreps D^k in the Wigner canonical form it is more convenient to split the row and column indices of $D^k(g)$ into pairs of indices, $\bar{a}a$, where $a = 1,...,\dim \Gamma^k$ and $\bar{a} = 1,2$ distinguishes the submatrices Γ^k in D^k for coirreps of type II and type III, and $\bar{a} = 1$ for type I [see Eq. (2.6)].

In this notation the matrix elements of $M^k(q)$ are of the form

$$M^k(q)_{m'\bar{a}'a',m\bar{a}a} = L^k(q)_{m'\bar{a}',m\bar{a}} \delta_{a'a}, \quad q \in Q^k. \tag{3.7}$$

Combining Eq. (3.5) with Eq. (3.6) we get

$$U(q)(qS^k)(e_k^R \times U^k_{(q)})^\dagger = S^k(L^k(q) \times I^k). \tag{3.8}$$

From this equation and Eq. (3.7) it is seen that the blocks S^k can be divided into p_k sub-blocks $S^{k,m}_{\bar{a}}$, each of them containing the columns $S^{k,m}_{\bar{a}a}$, $a = 1,...,\dim \Gamma^k$. These blocks, endowed with a scalar product

$$\langle S^{k,m}_{\bar{a}}, S^{k,m'}_{\bar{a}'} \rangle = \mathrm{trace}(S^{k,m}_{\bar{a}})^\dagger S^{k,m'}_{\bar{a}'} = \dim \Gamma^k \delta_{mm'} \delta_{\bar{a}\bar{a}'}, \tag{3.9}$$

form a unitary space. Taking into account (3.8) we can define the operators $T(q)$ acting in this space:

$$(T(q)S^k)^m_{\bar{a}} = U(q) \sum_{\bar{a}'} (qS^k)^m_{\bar{a}'} U^k(q)^\dagger_{\bar{a}'\bar{a}} = \sum_{m'\bar{a}'} S^{k,m'}_{\bar{a}'} L^k(q)_{m'\bar{a}',m\bar{a}}, \quad q \in Q^k. \tag{3.10}$$

Here $U^k(q)_{\bar{a}'\bar{a}}$ is the corresponding submatrix of $U^k(q)$.

It can be shown that the set of operators $T(q)$ defined by (3.10), $U(q) \in \bar{Q}$, and $U^k(q) \in \bar{Q}^k$, form an operator group $\tilde{Q}^k$. The operator $T(q) \in \tilde{Q}^k$ is unitary/antiunitary iff $q \in Q^k$ is unitary/antiunitary. The corresponding set of matrices $L^k(q)$ forms a corep of the group $\tilde{Q}^k$.

The further discussion concerns mainly the following two properties of the set of matrices $L^k(q)$: (i) they belong to the commutator algebra of the corep $(e_k^R \times D^k)$ of the group $G(H)$ [here we identify $a(L^k)$ with $a(M^k)$]; and (ii) they form a corep L^k of the operator group $\tilde{Q}^k$, Eq. (3.10).

First, we determine the structure of matrices $L^k(q)$ as elements of the commutator algebra for the three different types of coreps D^k.

For D^k of type I we have $\dim D^k = \dim \Gamma^k$ and

$$L^k(q)_{m'\bar{a}',m\bar{a}} = L^{k_1}(q)_{m'm} \delta_{\bar{a}'\bar{a},11}, \tag{3.11}$$

where L^{k_1} is an orthogonal matrix with $\dim L^{k_1} = m_k$. We should note that for the case of ordinary reps the commuting matrix L^k is of a similar type but in general it will be unitary instead of orthogonal.

For D^k of type II and type III, the matrices are of the form

$$L^{k_{\mathrm{II}}} = \begin{vmatrix} z_{11} & z_{12} & \cdots & z_{1m_k} \\ z_{21} & z_{22} & \cdots & z_{2m_k} \\ \cdots & \cdots & \cdots & \cdots \\ z_{mk1} & z_{mk2} & \cdots & z_{m_km_k} \end{vmatrix}, \quad \text{where } z_{ij} = \begin{vmatrix} x_{ij} & y_{ij} \\ -y^*_{ij} & x^*_{ij} \end{vmatrix}, \quad x_{ij}, y_{ij} \in \mathrm{C}; \tag{3.12}$$

$$L^{k_{\mathrm{III}}} = \begin{vmatrix} x_{11} & 0 & x_{12} & \cdots & x_{1m_k} & 0 \\ 0 & x_{11}^* & 0 & \cdots & 0 & x_{1m_k}^* \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & x_{m_k1}^* & 0 & \cdots & 0 & x_{m_km_k}^* \end{vmatrix}, \quad x_{ij} \in \mathbb{C}. \tag{3.13}$$

We can say something more about the $L^{k_{\mathrm{III}}}$ matrix. It is obvious that it can be reduced into a block-diagonal form

$$L^{k_{\mathrm{III}}} = \begin{vmatrix} L^{k_{\mathrm{III}},1} & 0 \\ 0 & L^{k_{\mathrm{III}},2} \end{vmatrix}, \tag{3.14}$$

where the blocks refer to the barred indices $\bar{a}'\bar{a}$ and

$$L^{k_{\mathrm{III}},1} = \begin{vmatrix} x_{11} & x_{12} & \cdots & x_{1m_K} \\ x_{21} & x_{22} & \cdots & x_{2m_k} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m_k1} & x_{m_k2} & \cdots & x_{m_km_k} \end{vmatrix},$$

$$\dim L^{k_{\mathrm{III}},1} = m_k, \tag{3.15a}$$

$$L^{k_{\mathrm{III}},2} = (L^{k_{\mathrm{III}},1})^*. \tag{3.15b}$$

$$L^{k_{\mathrm{III}}}(q)_{m'\bar{a}',m\bar{a}} = L^{k_{\mathrm{III}},1}(q)_{m'm}\delta_{\bar{a}'\bar{a},11} + L^{k_{\mathrm{III}},2}(q)_{m'm}\delta_{\bar{a}'\bar{a},22}. \tag{3.16}$$

The second aspect concerning the $L^k(q)$ matrices is related to the fact that the matrices $L^k(q)$, $q \in Q^k$, form a corep L^k of $\tilde{Q}^k$. For the type I coirreps D^k the $p_k = m_k$ linearly independent blocks $S^{k,m} = S_1^{k,m}$, endowed with the scalar product (2.9), span a carrier space of dimension m_k of the corep $L^{k_{\mathrm{I}}}$ of $\tilde{Q}^{k_{\mathrm{I}}}$:

$$(T(q)S^k)^m = \sum_{m'} L^{k_{\mathrm{I}}}(q)_{m'm} S^{k,m'}. \tag{3.17}$$

As we have already pointed out this result resembles very much the results for the case of ordinary reps[1] with the exception that for the case of coreps $L^{k_{\mathrm{I}}} = L^{k_{\mathrm{I}}*}$.

For the type II coreps D^k

$$(T(q)S^k)_{\bar{a}}^m = \sum_{m'\bar{a}'} L^{k_{\mathrm{II}}}(q)_{m'\bar{a}',m\bar{a}} S_{\bar{a}'}^{k,m'}, \tag{3.18}$$

where $L^{k_{\mathrm{II}}}$ is of the type (3.12). The difference between (3.18) and (3.17) [or the analogous relation for reps, Eq. (1.4) in I] is obvious. For the type I coreps D^k the matrix $L^{k_{\mathrm{I}}}$ realizes a mixing of the basis functions $S^{k_{\mathrm{I}},m}$ labeled by the different multiplicity indices m (m being the multiplicity label of D^k in R). For the type II coreps D^k the pair of sets $S_1^{k_{\mathrm{II}}}$ and $S_2^{k_{\mathrm{II}}}$ are grouped together into a double set

$$S^{k_{\mathrm{II}}} = \{S_{\bar{a}}^{k,m},\ m = 1,\ldots,m_k,\ \bar{a} = 1,2\}$$

and the matrix $L^{k_{\mathrm{II}}}$ intermixes all the functions belonging to $S^{k_{\mathrm{II}}}$. In other words, the functions from a double set span a carrier space of dimension $p_k = 2m_k$ of a corep of $\tilde{Q}^{k_{\mathrm{II}}}$.

For D^k of type III the relation (3.8) splits into two relations, because of (3.16),

$$(T(q)S^k)_1^m = \sum_{m'} L^{k_{\mathrm{III}},1}(q)_{m'm} S_1^{k,m'}, \tag{3.19a}$$

$$(T(q)S)_2^m = \sum_{m'} L^{k_{\mathrm{III}},2}(q)_{m'm} S_2^{k,m'},$$

$$L^{k_{\mathrm{III}},2}(q)_{m'm} = L^{k_{\mathrm{III}},1}(q)_{m'm}^*. \tag{3.19b}$$

Thus, the sets $S_1^{k_{\mathrm{III}}}$ and $S_2^{k_{\mathrm{III}}}$ span carrier spaces for two coreps of $\tilde{Q}^k$, $L^{k_{\mathrm{III}},1}$, and $L^{k_{\mathrm{III}},2}$, where $\dim L^{k_{\mathrm{III}},1} = \dim L^{k_{\mathrm{III}},2} = m_k$. The matrices $L^{k_{\mathrm{III}},\bar{a}}$, $\bar{a} = 1,2$, are related by (3.15b). If the complex conjugation is a symmetry operation for the coreps $L^{k_{\mathrm{III}},1}$, $L^{k_{\mathrm{III}},2}$, then $L^{k_{\mathrm{III}},1} \sim L^{k_{\mathrm{III}},2}$.

In order to take more advantage of the fact that the $L^k(q)$ matrices form a corep of $\tilde{Q}^k$ we need the concept of "G-equivalent" bases[18] ($G = \tilde{Q}^k$ in our case). It is always possible to choose the basis of a carrier space so that the matrices coincide identically with those of a "standard set of coreps" (which means a set of corep matrices chosen and fixed in a definite way). If there exist several possibilities to choose such a basis these bases are called "G-equivalent."

Considering the fact that the reducing matrix blocks $S_{\bar{a}}^{k,m}$ are the basis functions for a corep L^k of $\tilde{Q}^k$, it is natural to determine this basis in such a way that L^k is reduced to a block-diagonal form. In addition every irreducible constituent L^s is required to occur in standard form

$$L^k(q) = \bigoplus_s (e_s^k \times L^s(q)), \quad q \in Q^k. \tag{3.20}$$

Accordingly, the blocks $S_{\bar{a}}^{k,m}$ of S^k form a $\tilde{Q}^k$-equivalent basis.

To what extent does this reduction solve, eliminate, or reduce the arbitrariness of S^k? We should note that even if we fix the corep matrices $L^s(q)$ the corresponding basis is not uniquely determined. Due to the Schur lemma the basis is only fixed up to a unitary matrix M^s belonging to the commuting algebra $a(L^s)$ of the corresponding irreducible corep L^s. These matrices depend on the type of the corep and are of the following form:

$$\text{type I:} \quad M^s = \pm I^s; \tag{3.21a}$$

$$\text{type II:} \quad M^s = \begin{vmatrix} x & y \\ -y^* & x^* \end{vmatrix} \times I^s, \quad x,y \in \mathbb{C}, \quad |x|^2 + |y|^2 = 1; \tag{3.21b}$$

$$\text{type III:} \quad M^s = \begin{vmatrix} x & 0 \\ 0 & x^* \end{vmatrix} \times I^s, \quad x \in \mathbb{C}, \quad |x| = 1. \tag{3.21c}$$

Obviously for the special case $L^k = L^s$ the basis $S_{\bar{a}}^{k,m}$ is determined up to a sign, up to three real parameters, and up to one real parameter for corep types I, II, and III, respectively. This "inherent arbitrariness" in fact determines the highest level of uniqueness that can be achieved by means of the auxiliary operator method. We should note that for the case of ordinary reps the inherent arbitrariness is only of the phase factor type.

The procedure of solving the arbitrariness problem can be extended by including additional operations, e.g., permutations if R is a direct product corep.[3] However, this will not affect the inherent arbitrariness. One possibility to overcome the "Schur lemma barrier" (2.21) is to adopt some sensible conventions. For instance, this can be done by standardization of phases, based on the Racah lemma as discussed in details in Ref. 15.

If a given D^k occurs only once in the reduction of R, i.e., $m = m_k = 1$, it is necessary to determine the corresponding block $S_{\bar{a}}^{k,1}$ by standard methods (e.g., the projection method).

If $m_k > 1$ then the auxiliary group can be used as in I to reduce the multiplicity problem. If it is not completely resolved by this approach the remaining arbitrariness of the blocks can be eliminated by further convention, e.g., by means of the Schmidt orthonormalization procedure.

To exploit the auxiliary groups $Q^{\rm CO}$ to the utmost degree one has to generate the blocks $S_{\bar{a}}^{l,m}$, $l\in[k]$, from the already known blocks $S_{\bar{a}}^{k,m}$. This can be done using the generating relation

$$S_{\bar{a}}^{l,m} = U(q_l^{(k)})(q_l^{(k)} S_{\bar{a}}^{k,m}), \tag{3.22}$$

which is identical to Eq. (2.42) in I.

The following two examples illustrate the scheme proposed here.

IV. EXAMPLES

A. An eight-dimensional corep of the gray double point group $C_4^* \otimes \Theta$

As a first example we consider a reducing matrix for a corep of the antiunitary gray double point group $G(H) = C_4^* \otimes \Theta$ (see, e.g., Ref. 7). The unitary subgroup $H = C_4^*$ is isomorphic to $C_8 \simeq \langle C_{4z}^4 = \bar{E}, \bar{E}^2 = E\rangle$. We define the group

$$\mathrm{AUT} \simeq \mathrm{Aut}(C_8 \otimes C_2) \simeq D_2, \tag{4.1}$$

where we have taken into account that AUT $C_2 = C_1$ and the cyclic group C_8 has only outer automorphisms.

From the character table of $C_4^* \otimes \Theta$ (Table I) it is seen that the only nontrivial one-dimensional coirrep is D^2. In accordance with

$$aD(g) = D(g) \times D^2(g) \tag{4.2}$$

and the corep multiplication table (Table II) we find that the operator a generates the group

$$\mathrm{ASS} = \langle a\rangle \simeq C_2. \tag{4.3}$$

The group $Q^{\rm CO}$ has the following structure:

$$Q^{\rm CO} = C_2 \otimes (D_2 \otimes C_2') = \langle a, b_1, b_2, c\rangle. \tag{4.4}$$

TABLE I. Character table of $C_4^* \otimes \Theta$.

D^k	Γ^k	Type	E	C_4	C_4^2	C_4^3	θ
D^1	Γ^1	I	1	1	1	1	1
D^2	Γ^2	I	1	-1	1	-1	1
D^3	$\Gamma^3+\Gamma^4$	III	2	0	-2	0	0
D^5	$\Gamma^5+\Gamma^6$	III	2	$\sqrt{2}$	0	$\sqrt{2}$	0
D^8	$\Gamma^8+\Gamma^7$	III	2	$-\sqrt{2}$	0	$-\sqrt{2}$	0

TABLE II. Corep multiplication table of $C_4^* \otimes \Theta$.

$K \backslash k'$	1	2	3	5	8
1	1	2	3	5	8
2	2	1	3	8	5
3	3	3	1^2+2^2	$5+8$	$5+8$
5	5	8	$5+8$	1^2+3	2^2+3
8	8	5	$5+8$	2^2+3	1^2+3

We note that the semidirect product in (2.8) is reduced to a direct product, because (i) D^2 is a real corep, so $ac = ca$; and (ii) D^2 is the only nontrivial one-dimensional corep of $C_4^* \otimes \Theta$, i.e., $bD^2 = D^2$ and $ab = ba$.

The action of the generators of $Q^{\rm CO}$ on the classes of the equivalence coreps is shown in the qk table (Table III).

To illustrate our scheme we calculate the matrix S for the reducible eight-dimensional corep $R = D^3 \times D^8 \times D^8 \sim 2D^1 + 2D^2 + 2D^3$. Using the qk table the group Q is easily determined as

$$Q = Q^{\rm CO} = \langle a, b_1, b_2, c\rangle. \tag{4.5}$$

The Q-classes are therefore

$$[1] = \{D^1, D^2\}, \tag{4.6a}$$

$$[3] = \{D^3\}, \tag{4.6b}$$

$$[5] = \{D^5, D^8\}. \tag{4.6c}$$

The corresponding Q^k groups and coset representatives are the following:

$$Q^1 = \langle b_1, b_2, c\rangle \simeq D_2 \otimes \Theta, \quad q_1^{(1)} = e, \quad q_2^{(1)} = a, \tag{4.7a}$$

$$Q^3 = Q^{\rm CO} \simeq D_{2h} \otimes \Theta, \tag{4.7b}$$

$$Q^5 = \langle b_1, c\rangle \simeq C_2 \otimes \Theta, \quad q_1^{(5)} = e, \quad q_2^{(5)} = a, \quad q_3^{(5)} = b_2. \tag{4.7c}$$

The matrices $R(g)$ of the generators $g_1 = C_{4z}$ and $g_2 = \theta$ of $G(H)$ are

$$R(C_{4z}) = \mathrm{diag}(1, i, i, -1, -1, -i, -i, 1), \tag{4.8a}$$

$$R(\theta) = \text{skew diag}(1, -1, -1, 1, 1, -1, -1, 1). \tag{4.8b}$$

These matrices are obtained as Kronecker products of the corresponding matrices $D^k(g)$ taken from[7]

$$D^3(C_{4z}) = \mathrm{diag}(i, -i), \quad D^3(\theta) = \text{skew diag}(1,1),$$

$$D^8(C_{4z}) = \mathrm{diag}(-x^*, -x), \tag{4.9a}$$

$$x = \exp(i\pi/4), \quad D^8(\theta) = \text{skew diag}(-1,1). \tag{4.9b}$$

The nonzero elements of the matrices $U(q)$ for the generators of the group $Q \simeq Q^{\rm CO}$ are then

TABLE III. qk-table of $C_4^* \otimes \Theta$.

$qD^k \backslash D^k$	D^1	D^2	D^3	D^5	D^8
aD^k	D^2	D^1	D^3	D^8	D^5
b_1D^k	D^1	D^2	D^3	D^5	D^8
b_2D^k	D^1	D^2	D^3	D^8	D^5
cD^k	D^1	D^2	D^3	D^5	D^8

$$U(a):\ (1,5) = (2,6) = -(3,7) = (4,8) = (5,1)$$
$$= -(6,2) = -(7,3) = (8,4) = 1\,, \quad (4.10a)$$
$$U(b_1) = U(b_2) = U(c)$$
$$= \text{skew diag}(1,-1,-1,1,1,-1,-1,1)\,. \quad (4.10b)$$

This set of matrices satisfies the relations

$$(U(a))^2 = (U(c))^2 = E\,, \quad (4.11a)$$
$$U(a)U(c) = U(c)U(a)\,, \quad (4.11b)$$

and forms the group $\overline{Q} \simeq C_2 \otimes \Theta$.

The next step in our procedure is the determination of the groups $\overline{Q}^k = \{U(q)\,|\,q\in Q^k\}$ for the coreps D^k occurring in the decomposition of R.

Using the matrices D^k given in (4.9) and Eqs. (2.23), we get the following matrices $U^k(q)$.

For the corep D^1 of class [1], as $\dim D^1 = 1$, we can choose

$$U^1(q) = 1, \quad \text{for all } q\in Q^1, \quad \text{hence } \overline{Q}^1 \simeq C_1\,. \quad (4.12a)$$

For the corep D^3 of the class [3],

$$U^3(a) = U^3(c) = U^3(b_1) = U^3(b_2)$$
$$= \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}, \text{ hence } \overline{Q}^3 \simeq C_2\,. \quad (4.12b)$$

Now we are in the position to determine the general form of blocks of S. From the theory of coreps, especially the Schur lemma for reducible coreps, it follows that the reducing matrix block S^1, which satisfies

$$R(C_{4z})S^1 = S^1D^1(C_{4z})\,, \quad (4.13a)$$
$$R(\theta)S^{1*} = S^1D^1(\theta)\,, \quad (4.13b)$$

is determined up to two-dimensional orthogonal transformation.

Using (4.8) and (4.13) we obtain, for S^1,

$$S^1 = \begin{vmatrix} a+ib \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ a-ib \end{vmatrix}, \quad \text{where } a,b\in R\,. \quad (4.14)$$

The action of the operators $T(q)$, $q = b_1b_2, c\in Q^1$, is given by the following relations:

$$T(b_1)S^1 = U(b_1)S^1U^1(b_1)^\dagger = S^{1*}\,, \quad (4.15a)$$
$$T(b_2)S^1 = U(b_2)S^1U^1(b_2)^\dagger = S^{1*}\,, \quad (4.15b)$$
$$T(c)S^1 = U(c)S^{1*}U^1(c)^\dagger = S^1\,. \quad (4.15c)$$

So $\widetilde{Q}^1 \simeq C_2 \otimes \Theta$, and the character table of this group is given in Table IV. Obviously the multiplicity problem is solved because if the reducing matrix blocks $S^{1,1}$ and $S^{1,2}$ are fixed by

$$S^{1,1}:\ a = 1, \quad b = 0\,, \quad (4.16a)$$
$$S^{1,2}:\ a = 0, \quad b = 1\,, \quad (4.16b)$$

they turn out to be basis functions for the two inequivalent one-dimensional coirreps $L^1 = D^1$ and $L^2 = D^2$. This is seen from the corresponding $L^1(q)$ matrices

$$L^1(e) = \text{diag}(1,1)\,, \quad (4.17a)$$
$$L^1(b) = \text{diag}(1,-1)\,, \quad (4.17b)$$
$$L^1(c) = \text{diag}(1,1)\,. \quad (4.17c)$$

TABLE IV. Character table of $C_2 \otimes \Theta$.

D^k	Γ^k	Type	E	C_2	Θ	$C_2\Theta$
D^1	Γ^1	I	1	1	1	1
D^2	Γ^2	I	1	-1	1	-1

As the corep D^2 of $G(H) = C_4^* \otimes \Theta$ belongs to the same class [Eq. (4.6a)] we can use Eqs. (3.22) and (4.7a) and the matrix $U(q_2^{(1)}) = U(a)$, Eq. (4.10a), to find the block $S^{2,m}$.

The general form of $S^{3,m}$ can be obtained from the defining relations

$$R(C_{4z})S^3 = S^3D^3(C_{4z})\,, \quad (4.18a)$$
$$R(\theta)S^{3*} = S^3D^3(\theta)\,, \quad (4.18b)$$

where $R(g)$ and $D^3(g)$ are given by (4.8) and (4.9), respectively. Since D^3 is a type III coirrep it follows from the Schur lemma that $S^{3,m}$ is determined up to two complex parameters (3.13). For the general form of the blocks $S^{3,m}$ we get

$$S^{3,m} = \begin{vmatrix} 0 & 0 \\ a_1+ib_1 & 0 \\ a_2+ib_2 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & -a_2+ib_2 \\ 0 & -a_1+ib_1 \\ 0 & 0 \end{vmatrix}, \quad \text{where } a_i,b_i\in R\,. \quad (4.19)$$

The action of the operators $T(a)$, $T(b)$, $T(c)$ on the $S^{3,m}$ blocks (3.10) shows that $\widetilde{Q}^3 \simeq D_2 \otimes \Theta$. The character table of this group is given in Table V. If we choose the blocks $S^{3,1}$ and $S^{3,2}$ in the form

$$S^{3,1}:\ a_1 = a_2 = 1\,, \quad b_1 = b_2 = 0\,, \quad (4.20a)$$
$$S^{3,2}:\ a_1 = a_2 = 0\,, \quad b_1 = -b_2 = 1\,, \quad (4.20b)$$

TABLE V. Character table of $D_2 \otimes \Theta$.

D^k	Γ^k	Type	E	C_{2z}	C_2'	C_2''	θ
D^1	Γ^1	I	1	1	1	1	1
D^2	Γ^2	I	1	-1	1	-1	1
D^3	Γ^3	I	1	1	-1	-1	1
D^4	Γ^4	I	1	-1	-1	1	1

the set $S^3_{11} = \{S^{3,1}_{11}, S^{3,2}_{11}\}$ spans a carrier space for the reducible corep of $\widetilde{Q}^3$, Eq. (3.19a), given by the matrices $L^{3,1}(q)$:

$$L^{3,1}(e) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}, \quad L^{3,1}(a) = \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix}, \quad L^{3,1}(b) = \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix}, \quad L^{3,1}(c) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}. \tag{4.21}$$

So $S^{3,1}_{1}$ is transforming according to the identity coirrep, while $S^{3,2}_{11}$ is a basis function of one of the nontrivial one-dimensional coirreps [e.g., D^4 if $T(a) \leftrightarrow C_{2z}$, $T(b) \leftrightarrow C'_2$].

In this way we succeed in solving completely the multiplicity problem for the reducing matrix for $R = D^3 \times D^8 \times D^8$. The final form of S reads

$$S = \begin{array}{c|cccccccc} kma & 111 & 121 & 211 & 221 & 311 & 312 & 321 & 322 \\ \hline & \sqrt{1/2} & i\sqrt{1/2} & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & \sqrt{1/2} & 0 & i\sqrt{1/2} & 0 \\ & 0 & 0 & 0 & 0 & \sqrt{1/2} & 0 & -i\sqrt{1/2} & 0 \\ & 0 & 0 & \sqrt{1/2} & i\sqrt{1/2} & 0 & 0 & 0 & 0 \\ & 0 & 0 & \sqrt{1/2} & -i\sqrt{1/2} & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 & -\sqrt{1/2} & 0 & i\sqrt{1/2} \\ & 0 & 0 & 0 & 0 & 0 & -\sqrt{1/2} & 0 & -i\sqrt{1/2} \\ & \sqrt{1/2} & -i\sqrt{1/2} & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \tag{4.22}$$

B. The matrix of Clebsch–Gordan coefficients of $D^4 \times D^4$ of the black and white double point group $O^*(T^*)$

Next we discuss the antiunitary black and white double point group $O^*(T^*)$ and a reducing matrix for the Kronecker product $D^4 \times D^4$, composed of Clebsch–Gordan coefficients. In the notations of Bradley and Cracknell[7]

$$O^*(T^*) = T^* + \theta C_{2a} T^*, \tag{4.23}$$

where the asterisk indicates that we are dealing with the double groups.

As a set of generators for $O^*(T^*)$ we choose for the unitary subgroup $T^* \lhd O^*(T^*)$ the generators $(C_{2z})^2 = \overline{E}$ and $(C^+_{31})^3 = \overline{E}$ and add the fixed antiunitary coset representative $a_0 = \theta C_{2a}$. Because of the abstract isomorphism $O^*(T^*) \simeq O^*$ (where $\theta C_{2a} T \leftrightarrow C_{2a} T$), the defining relations and the group multiplication table of $O^*(T^*)$ are the same as for the double octahedral group O*.

The table of corep characters is determined by the irrep table of the unitary subgroup T^* since all the coreps are of the type I (see Table VI). For convenience we also give the corep multiplication table of $O^*(T^*)$ (Table VII). It is seen from Table VI, that there exist three one-dimensional coreps of $O^*(T^*)$, viz. D^1, D^2, D^3. They belong to one class of associated coreps $\{D^1, D^2, D^3\}$ (see Ref. 15), i.e., $(D^k)^3 = D^1$ and $(D^k)^2 = D^{k'}$, $k \neq k' = 2,3$. In accordance with the definition (1.9) the group ASS is generated by association with D^3, and the corresponding operator a determines the group

$$\text{ASS} = \langle a \rangle \simeq C_3 . \tag{4.24}$$

The action of the operator $c \in \text{CON} \simeq C_2$ is obvious from Table VI.

TABLE VI. Character table of $O^*(T^*)$. Here $\omega = \exp(i\pi/3)$.

D^k	Γ^k	Type	E	C_{2z} $\overline{C}_{2z}$	C^+_{31}	C^-_{31}	θC_{2a} $\theta\overline{C}_{2a}$
D^1	Γ^1	I	1	1	1	1	1
D^2	Γ^2	I	1	1	ω^2	ω^4	1
D^3	Γ^3	I	1	1	ω^4	ω^2	1
D^4	Γ^4	I	3	-1	0	0	1
D^5	Γ^5	I	2	0	1	1	$\sqrt{2}$
D^6	Γ^6	I	2	0	ω^4	ω^2	$-\sqrt{2}$
D^7	Γ^7	I	2	0	ω^2	ω^4	$-\sqrt{2}$

The determination of the group AUT is facilitated by the fact that the element $\overline{E}$ belongs to the center of $O^*(T^*)$, i.e., $\text{Aut}\, O^* \simeq \text{Aut}\, O \simeq O$. So the only automorphism of $O^*(T^*)$ that is outer for the unitary subgroup T^*, is of second order, and can be realized by a conjugation with $\theta C_{2b} \in O^*(T^*)$ is

$$\beta(C_{2z}) = \overline{C}_{2z}, \quad \beta(C^+_{31}) = C^-_{31}, \quad \beta(\theta C_{2a}) = \theta\overline{C}_{2a} . \tag{4.25}$$

Consequently

$$\text{AUT} = \langle b \rangle \simeq C_2 . \tag{4.26}$$

Hence we obtain the full group

$$Q^{\text{CO}} = C_3 \circledS (C_2 \otimes C'_2) = \langle a;b,c \rangle \simeq D_6(D_3) . \tag{4.27}$$

From the definitions of $q \in Q^{\text{CO}}$ and Table VI and Table VII we construct the qk table of $O^*(T^*)$ (see Table VIII).

The present example deals with the construction of the reducing matrix S, which carries out the transformation of the reducible corep $R = D^4 \times D^4$ into a direct sum of its irreducible constituents:

$$S^\dagger R(g) S^{(g)} = (D^1 \oplus D^2 \oplus D^3 \oplus D^4)(g) . \tag{4.28}$$

Since the corep R is a Kronecker product the elements of the reducing matrix S are the familiar Clebsch–Gordan coefficients.

From the gk table it follows that

$$Q = Q^{\text{CO}} = \langle a,b,c \rangle \simeq D_6(D_3) . \tag{4.29}$$

We can easily construct the Q-classes

TABLE VII. Corep multiplication table of $O^*(T^*)$.

	1	2	3	4	5	6	7
1	1	2	3	4	5	6	7
2	2	3	1	4	6	7	5
3	3	1	2	4	7	5	6
4	4	4	4	1+2+3+4+4	5+6+7	5+6+7	5+6+7
5	5	6	7	5+6+7	1+4	2+4	3+4
6	6	7	5	5+6+7	2+4	3+4	1+4
7	7	5	6	5+6+7	3+4	1+4	2+4

$$[1] = \{D^1,D^2,D^3\},\quad [4] = \{D^4\},\quad [5] = \{D^5,D^6,D^7\}. \tag{4.30}$$

The corresponding Q^k-groups and coset representatives are

$$\begin{aligned} &Q^1 = \langle a,c\rangle \simeq C_2 \otimes \Theta, \\ &\quad q_1^{(1)} = e,\quad q_2^{(1)} = a,\quad q_3^{(1)} = a^2; \\ &Q^4 = Q^{\rm CO}; \\ &Q^5 = \langle b,c\rangle \simeq C_2 \otimes \Theta, \\ &\quad q_1^{(5)} = e,\quad q_2^{(5)} = a,\quad q_3^{(5)} = a^2. \end{aligned} \tag{4.31}$$

The matrices $R(g) = D^4(g) \times D^4(g)$ for the generators $C_{2z}, C_{31}^+, \theta C_{2a}$ are obtained by using the following standard corep matrices[7] of D^4:

$$\begin{aligned} D^4(C_{2z}) &= \begin{vmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{vmatrix}, \\ D^4(C_{31}^+) &= \begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & i \\ -i & 0 & 0 \end{vmatrix}, \\ D^4(\theta C_{2a}) &= \begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix}. \end{aligned} \tag{4.32}$$

For the unitary matrices $U(q)$, $q = a,b,c \in Q$ [Eq. (2.21)], we obtain

$$U(a) = \mathrm{diag}(\omega,\omega,\omega,\omega^*,\omega^*,\omega^*,-1,-1,-1), \tag{4.33a}$$

$$U(c) = \mathrm{diag}(1,1,-1,1,1,-1,-1,-1,1). \tag{4.33b}$$

$$U(b)_{ij} = \begin{cases} 1, & ij = 15,24,36,42,51,63,78,87,95, \\ 0, & \text{otherwise}. \end{cases} \tag{4.33c}$$

TABLE VIII. The qk table of $O^*(T^*)$.

qD^k \ D^k	D^1	D^2	D^3	D^4	D^5	D^6	D^7
a^2D^k	D^2	D^3	D^1	D^4	D^6	D^7	D^5
$a\,D^k$	D^3	D^1	D^2	D^4	D^7	D^5	D^6
$b\,D^k$	D^1	D^3	D^2	D^4	D^5	D^7	D^6
$c\,D^k$	D^1	D^3	D^2	D^4	D^5	D^7	D^6

To determine the $\overline{Q}$ group we have to consider the following generating relations satisfied by the $U(q)$ matrices:

$$\begin{aligned} &(U(a))^6 = (U(b))^2 = (U(c))^2 = E, \\ &U(b)U(c) = U(c)U(b), \\ &U(a)U(c) = U(c)(U(a))^5, \\ &U(a)U(b) = U(b)(U(a))^5. \end{aligned} \tag{4.34}$$

From the relations (4.34) we find that $\overline{Q} \simeq C_6 \circledS (C_2 \otimes C_2')$.

The determination of $U^1(q)$ is trivial because $\dim D^1 = 1$:

$$U^1(q) = 1,\quad q = b,c \in Q^1,\quad \text{hence } \overline{Q}{}^1 \simeq C_1 \times \Theta. \tag{4.35}$$

Because of $Q^4 = Q^{\rm CO}$, we need $U^4(q)$ for $q \in Q^{\rm CO}$. Taking into account the corep matrices of D^4 [Eq. (4.32)], we obtain the following $U^4(q)$ matrices:

$$\begin{aligned} U^4(a) &= \begin{vmatrix} \omega & 0 & 0 \\ 0 & \omega^* & 0 \\ 0 & 0 & -1 \end{vmatrix}, \\ U^4(b) &= \begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix}, \\ U^4(c) &= \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{vmatrix}. \end{aligned} \tag{4.36}$$

The above matrices $U^4(q)$, $q \in Q^4$, satisfy the same defining relations (4.34), as the matrices $U(q)$ of $\overline{Q}$, whence $\overline{Q}{}^4 \simeq \overline{Q}$.

Now we are ready for the determination of blocks $S^{k,m}$ of S. The use of the relations of the type (2.1) for the generators of $O^*(T^*)$ suffices to obtain the general form of S^1.

$$R(g)S^{1(g)} = S^1 D^1(g),\quad g = C_{2z}, C_{31}^+, \theta C_{2a}. \tag{4.37}$$

This block is determined up to one real parameter since D^1 is a type I corep. The block S^1 is a column matrix with nonzero elements $d_{11} = d_{51} = -d_{91} = d$. To obtain a unitary S we put $d = \sqrt{1/3}$. Now S^2 and S^3 are obtained from the relations (3.22), (4.31), and (4.33a),

$$S^2 = U(a^2)S^1, \tag{4.38a}$$

$$S^3 = U(a)S^1. \tag{4.38b}$$

Proceeding in the same way as before we get the following general form of $S^{4,m}$:

$$S^{4,m} = \begin{vmatrix} 0 & 0 & 0 \\ 0 & 0 & -z \\ 0 & z^* & 0 \\ 0 & 0 & -z^* \\ 0 & 0 & 0 \\ z & 0 & 0 \\ 0 & z & 0 \\ z^* & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix}, \quad z = a + ib, \quad z \in \mathbb{C}. \tag{4.39}$$

In accordance with the consequences of the Schur lemma for coreps, $S^{4,m}$ is determined up to two real parameters.

The action of the operators $T(q)$ on the $S^{4,m}$ blocks is given by Eq. (3.10), where $U(q)$, $U^4(q)$, are taken from (4.33) and (4.36):

$$\begin{aligned} &T(b)S^4 = S^4, \\ &T(c)S^4 = -S^4, \\ &(T(abc))^6 S^4 = S^4, \\ &T(b)T(c) = T(c)T(b), \\ &T(abc)T(b) = T(b)(T(abc))^5. \end{aligned} \tag{4.40}$$

From these generating relations we conclude that the group $\widetilde{Q}^4$ is isomorphic to $D_6(D_3)$. [We obtain an antiunitary operator $T(abc)$ of sixth order that can be considered as the colored rotation θC_{6z}. Its square is a unitary operator corresponding to C_{3z}. The unitary operator $T(b)$ corresponds to C_{2x}; in this scheme $T(c) = C_{2z}\theta C_{2x}$.] The characters of $D_6(D_3)$ are given in Table IX (see Ref. 13).

To obtain the explicit form of the corep matrices of $\widetilde{Q}^4$ we have to make a choice for the parameters of $S^{4,m}$, i.e., we have to choose a basis defining the corep of $D_6(D_3)$. The simplest choice of $S^{4,1}$ and $S^{4,2}$ is determined by the following values of the parameters a,b:

$$\begin{aligned} &S^{4,1}: \quad a = 1, \quad b = 0, \\ &S^{4,2}: \quad a = 0, \quad b = 1. \end{aligned} \tag{4.41}$$

This gives rise to the following corep matrices $L^4(q) = D^3(q)$ of $D_6(D_3)$:

TABLE IX. Character table of $D_6(D_3)$.

D^k	Γ^k	Type	E	C_{3z}	C_2'	θC_2
D^1	Γ^1	I	1	1	1	1
D^2	Γ^2	I	1	1	-1	1
D^3	Γ^3	I	2	-1	0	0

$$\begin{aligned} L^4(a) &= \begin{vmatrix} -\frac{1}{2} & -\sqrt{3}/2 \\ \sqrt{3}/2 & -\frac{1}{2} \end{vmatrix}, \\ L^4(b) &= \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}, \\ L^4(c) &= \begin{vmatrix} -1 & 0 \\ 0 & 1 \end{vmatrix}. \end{aligned} \tag{4.42}$$

It is evident from (4.42) that this choice of the parameters a and b results in a Clebsch–Gordan coefficient matrix $S^{4,m}$, which is "self-consistent" only under complex conjugation and automorphism. That is, only complex conjugation and automorphism lead to symmetry operations for the Clebsch–Gordan coefficients of the "simple phase factor" type.[6] Here this phase factor reduces to a mere sign.

That it is not possible to obtain "simple phase factor" symmetries of the Clebsch–Gordan coefficients for all the three operations simultaneously follows from the irreducibility of the corep $D^3 = L^4(q)$. One may ask whether it is possible to diagonalize $L^4(a)$ by a suitable choice of the $S^{4,m}$ matrix. But since $D^4 + D^4$ is a type I corep, all the $L^4(q)$ matrices should be (real) orthogonal according to (3.10). But the diagonalization of $L^4(a)$ transforms it into a complex matrix that does not belong to the commutator algebra of the corep $D^4 + D^4$ of $O^*(T^*)$. Therefore, it is not possible to find a S^4-matrix, which is "self-consistent" under association. It should be noted that in the case of unitary irreps of T^* it is possible to diagonalize the associations because the corresponding commutator algebra is complex (see example A in I).

For the choice (4.41) of the parameters we obtain the following S matrix:

S \ k		1	2	3	4	4	4	4	4	4
m		1	1	1	1	1	1	2	2	2
a		1	1	1	1	2	3	1	2	3
$k'a'$	$k''a''$									
41	41	$\sqrt{1/3}$	$\sqrt{1/3}\omega^*$	$\sqrt{1/3}\omega$	0	0	0	0	0	0
41	42	0	0	0	0	0	$-\sqrt{1/2}$	0	0	$-i\sqrt{1/2}$
41	43	0	0	0	0	$\sqrt{1/2}$	0	0	$-i\sqrt{1/2}$	0
42	41	0	0	0	0	0	$-\sqrt{1/2}$	0	0	$i\sqrt{1/2}$
42	42	$\sqrt{1/3}$	$\sqrt{1/3}\omega$	$\sqrt{1/3}\omega^*$	0	0	0	0	0	0
42	43	0	0	0	$\sqrt{1/2}$	0	0	$\sqrt{1/2}$	0	0
43	41	0	0	0	0	$\sqrt{1/2}$	0	0	$i\sqrt{1/2}$	0
43	42	0	0	0	$\sqrt{1/2}$	0	0	$-i\sqrt{1/2}$	0	0
43	43	$-\sqrt{1/3}$	$\sqrt{1/3}$	$\sqrt{1/3}$	0	0	0	0	0	0 .

(4.43)

The matrix elements (4.43) coincide with the Clebsch–Gordan coefficients for coreps reducing the direct product $D^4 \times D^4$ of $O^*(T^*)$ as given in the full set of tables for the cubic groups.[13]

An additional restriction of the form of $S^{4,m}$ blocks follows from the requirement that $L^4(q)$ should form a "standard" corep of $\widetilde{Q}^4$ [see Eq. (3.20)]. It is irreducible and is of the first type, so the matrix S^4 is determined up to its "inherent arbitrariness," i.e., a sign [see Eq. (3.21a)].

If we want "the standard form" of $L^4(q) = D^3(q)$ of $D_6(D_3)$ following[13]

$$\begin{aligned} L^4(a) &= \begin{vmatrix} -\frac{1}{2} & \sqrt{3}/2 \\ \sqrt{3}/2 & -\frac{1}{2} \end{vmatrix}, \\ L^4(b) &= \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}, \\ L^4(c) &= \begin{vmatrix} 0 & -1 \\ -1 & 0 \end{vmatrix}, \end{aligned} \tag{4.44}$$

then we must take the blocks $S^{4,m}$ with parameters

$$\begin{aligned} S^{4,1}&: \ a = \sqrt{\tfrac{1}{2}}, \quad b = \sqrt{\tfrac{1}{2}}, \\ S^{4,2}&: \ a = \sqrt{\tfrac{1}{2}}, \quad b = -\sqrt{\tfrac{1}{2}}. \end{aligned} \tag{4.45}$$

The new $S^{4,m}$ blocks are related to (4.41) by the an orthogonal transformation. However, it is seen from (4.4) that they have not the symmetry of the "simple phase factor" type under automorphism and complex conjugation.

ACKNOWLEDGMENTS

The authors thank D. Allexandrova for helpful discussions.

This work is part of the Scientific–Technical Cooperation between Austria and Bulgaria. Financial support of research visits in Sofia (R. D. and P. K.) and in Vienna (M. I. A. and J. N. K.) is gratefully acknowledged.

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

[2] R. Dirl, P. Kasperkovitz, M. I. Aroyo, J. N. Kotzev, and M. N. Angelova-Tjurkedjieva, in *Group Theoretical Methods in Physics, Proceedings of the Third International Seminar*, Yourmala, USSR, May 1985, edited by M. A. Markov (Nauka, Moscow, 1986).

[3] R. Dirl, P. Kasperkovitz, M. I. Aroyo, J. N. Kotzev, and B. L. Davies, J. Phys. A **79**, L303 (1986).

[4] E. P. Wigner, *Group Theory* (Academic, New York, 1959).

[5] L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics* (Addison–Wesley, Reading, MA, 1981); *Racah–Wigner Algebra in Quantum Theory* (Addison–Wesley, Reading, MA, 1981); A. P. Jucis, J. B. Levinson, and V. V. Vanagas, *Mathematical Theory of Angular Momentum* (Mintis, Vilnyus, 1960).

[6] P. H. Butler, Philos. Trans. R. Soc. London Ser. A **277**, 545 (1975).

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

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

[9] J. O. Dimmock, J. Math. Phys. **4**, 1304 (1963).

[10] J. N. Kotzev, *On the Theory of Corepresentations of Magnetic Groups* (IRE AN USSR, Kharkov, 1972); J. N. Kotzev, Sov. Phys. Crystallogr. **19**, 286 (1974).

[11] R. Dirl, J. Math. Phys. **21**, 961, 968, 975, 983, 989, 997, 2011, 2015 (1980).

[12] P. Kasperkovitz, J. Math. Phys. **24**, 8 (1983); Physica A **114**, 54 (1982).

[13] J. N. Kotzev and M. I. Aroyo, Comm. Joint Inst. for Nuclear Res., P17-10987, Dubna, 1977; J. Phys. A **13**, 2275 (1980); **14**, 1543 (1981); **15**, 711, 725 (1981); J. N. Kotzev, M. I. Aroyo, and M. N. Angelova, Physica A **114**, 533 (1982); R. Comm. Soc. Edinburgh **19**, 253 (1983); Int. J. Quant. Chem. **27**, 45 (1985); J. Mol. Structure **115**, 123 (1984).

[14] J. N. Kotzev and M. I. Aroyo, in *Group Theoretical Methods in Physics, Proceedings of the International Seminar*, Zvenigorod, USSR, 1982, edited by M. A. Markov (Nauka, Moscow, 1983), Vol. I, p. 363; M. I. Aroyo and J. N. Kotzev, in *XIIIth International Colloquium on Group Theoretical Methods in Physics*, College Park, Maryland, 1984, edited by W. W. Zachary (World Scientific, Singapore, 1984), p. 360.

[15] J. N. Kotzev and M. I. Aroyo, in *Group Theoretical Methods in Physics, Proceedings of the XIth International Colloquium*, Istanbul, 1982, *Lecture Notes in Physics*, Vol. 180, edited by M. Serdaroglu and E. Inönü (Springer, Berlin, 1983), p. 332; J. Phys. A **17**, 727 (1984).

[16] P. M. van den Broek, J. Math. Phys. **20**, 2028 (1979).

[17] J. D. Newmarch and R. M. Golding, J. Math. Phys. **23**, 685 (1982).

[18] R. N. Haase and P. H. Butler, J. Phys. A **17**, 47 (1984).

Analysis on generalized superspaces

Yuji Kobayashi
Department of Mathematics, Tokushima University, Tokushima 770, Japan
Shigeaki Nagamachi
Technical College, Tokushima University, Tokushima 770, Japan

(Received 26 August 1985; accepted for publication 7 May 1986)

The analysis over σ-commutative algebras (generalized supercommutative algebras), that is, differentiation and integration for functions defined on superspace over a σ-commutative algebra, is studied.

I. INTRODUCTION

In the last ten years or so, the supersymmetric quantum field theory has been studied extensively,[1] and recently, attempts to give a mathematical foundation are very active,[2] where Grassmann algebras (supercommutative algebras) play an essential role. Supercommutativity is a generalization of commutativity but not the most general one. The generalized supercommutative algebras of Scheunert,[3] which we call σ-commutative algebras, are considered to be the most general. In our previous paper,[4] we developed the theory of matrices whose entries are elements of a σ-commutative algebra and studied Lie groups consisting of those matrices. These Lie groups are considered to be transformation groups of superspaces over σ-commutative algebras. Several authors studied the theory of infinitesimals of the Lie groups, that is, σ-Lie algebras (representation theory,[5] cohomology theory[6]).

In the present paper, we study analysis over σ-commutative algebras, that is, differentiation and integration of functions defined on a superspace over a σ-commutative algebra. Since the σ-commutative algebras include para-Bose and para-Fermi number systems as well as Grassmann algebras, our theory will provide a foundation to the study of commutation relations that appear in quantum field theories,[7] and will also help us towards deeper understanding of supersymmetry.

In the supersymmetric case, parameters of superspace are taken from a Grassmann algebra, but in our general case they are taken from the tensor product of the crossed product on the group of even grades and the generalized Grassmann algebra over odd generators. We define G^r-functions (r-times continuously differentiable functions) along the method of Rogers.[8] If r is greater than the nilpotency of superspace, then G^r-functions have standard forms called the standard expansions. Though the derivatives with respect to odd variables are not uniquely determined, we can choose the canonical one using the standard expansion. For a G^r-function f, the integration is defined as follows; first pick up the top $f_{1\cdots q}(x)$ from the standard expansion

$$f(x^1,\ldots,x^p,\xi^1,\ldots,\xi^q)=\sum_{r<q} f_{j_1\cdots j_r}(x)\xi^{j_1}\cdots\xi^{j_r},$$

and integrate it on the body $b(X)$ of the superspace X. The main theorem of this paper is a consistency theorem (Theorem 7.4), which states that under a change of variables the (super) Jacobian appears to adjust values of integrals. Recently, Rogers[9] formulated integration with respect to even variables as the contour integration in the supersymmetric case. In our general case, the same formulation is also possible, but it is practical and essential that integration with respect to even variables is taken to be the integration on the body as usual.[10]

This paper is organized as follows: In Sec. II, we define supernumbers, superspaces, and functions on superspaces (superfields). In Sec. III, we define the derivatives and the higher derivatives of superfields, and prove the analog of Taylor's expansion theorem (Proposition 3.8). In Sec. IV, we investigate functions defined on the body. To such functions correspond a set of ordinary functions. Using this correspondence, we define integration and prove its consistency under a change of variables on the body (Proposition 4.6). In Sec. V, we define the Jacobian matrix as usual, and the Jacobian to be the superdeterminant of the Jacobian matrix. The inverse mapping theorem (Theorem 5.4) is proved. In Sec. VI, we give the standard expansion for G^r-functions (not only G^∞-functions), where the concept of excessively C^r-functions is introduced. In Sec. VII, consistency of integration (Theorem 7.4) is proved in an elementary way, by decomposing the general form of change of variables into elementary ones. We give some remarks in the final section.

II. SUPERNUMBERS AND SUPERSPACES

In this section we prepare the basic notions, supernumber, and superspace on which we develop the (super) calculus.

First, we summarize some notions and notations introduced in the previous paper[4] that will be used in this paper.

Let G be a finite Abelian group and $\mathbb{F}$ be the real or the complex number field. We call a mapping $\sigma: G\times G\rightarrow\mathbb{F}$ a sign of G if it satisfies

(i) $\sigma(\alpha+\beta,\gamma)=\sigma(\alpha,\gamma)\sigma(\beta,\gamma)$,

(ii) $\sigma(\alpha,\beta)\sigma(\beta,\alpha)=1$,

for any $\alpha,\beta,\gamma\in G$, and the pair (G,σ) is called a signed group. By (ii) we see $\sigma(\alpha,\alpha)=\pm 1$ for any $\alpha\in G$. The event part $\{\alpha\in G\,|\,\sigma(\alpha,\alpha)=1\}$ and the odd part $\{\alpha\in G\,|\,\sigma(\alpha,\alpha)=-1\}$ of G are denoted by G_0 and G_1, respectively. Since $\sigma|_{G_0}$ is an even sign, there is a factor system $\phi: G_0\times G_0\rightarrow\mathbb{F}-\{0\}$ associated with $\sigma|_{G_0}$, that is, ϕ satisfies

 0022-2488/86/092247-10$02.50

(i) $\phi(\alpha,\beta+\gamma)\phi(\beta,\gamma)=\phi(\alpha,\beta)\phi(\alpha+\beta,\gamma)$,

(ii) $\phi(0,0)=1$,

(iii) $\phi(\alpha,\beta)/\phi(\beta,\alpha)=\sigma(\alpha,\beta)$,

for $\alpha,\beta,\gamma\in G_0$. We can choose ϕ so that $|\phi(\alpha,\beta)|=1$ (Proposition 2.7 of Ref. 4). Let $C=\oplus_{\alpha\in G_0}C_\alpha$ be the crossed product of F and G_0 defined by means of ϕ, that is, $C_\alpha=\mathrm{F}\cdot u_\alpha$ is the one-dimensional vector space over F with generator u_α of grade α and the multiplication in C is defined by

$$u_\alpha\cdot u_\beta=\phi(\alpha,\beta)u_{\alpha+\beta}.$$

Then C is a σ-commutative algebra over F by (iii) above.

A finite set I is called a G-set if it is linearly ordered and a grade $g(i)\in G$ is assigned to every element $i\in I$. Let L be an odd G-set, that is, each element l of L has an odd grade, and let V be the G-graded vector space over F with basis $\{v_l|l\in L\}$, where the grade of v_l is $g(l)$. Let B be the σ-Grassmann algebra over V defined by $B=T(V)/K$, where K is the ideal of the tensor algebra $T(V)$ over V generated by the elements $v_iv_j-\sigma(g(i),g(j))v_jv_i$ $(i,j\in L)$. Let M be a subset of L. M is a G-set in a natural way. The ordered product $\Pi_{l\in M}v_l$ is written as v_M. Then B is a G-graded σ-commutative algebra with a linear basis $\{v_M|M\subset L\}$ over F. Let $A=C\otimes_{\mathrm{F}}B$ be the graded tensor product of C and B over F, then A is a finite-dimensional σ-commutative algebra. In the rest of the paper we fix the algebra A of which elements are called *supernumbers*.

Any element a of A is expressed uniquely as

$$a=\sum_{\alpha,M}a_{\alpha,M}u_\alpha\otimes v_M, \tag{2.1}$$

where $a_{\alpha,M}\in\mathrm{F}$ and α ranges over the elements of G_0 and M the subsets of L. The norm $\|a\|$ of the element a is defined by

$$\|a\|=\sum_{\alpha,M}|a_{\alpha,M}|.$$

Then A is a Banach algebra over F.

The Grassmann–Banach algebra $B_{|L|}$ of Rogers[8] is obtained as a special case of our algebras, when $G=\mathbf{Z}_2$ and $\sigma(\alpha,\beta)=(-1)^{\alpha\beta}$, for $\alpha,\beta\in\mathbf{Z}_2$. In this case $G_0=\{0\}$ and the factor system ϕ is trivial and so C is equal to the base field F. On the other hand, B is the Grassmann algebra generated by $|L|$ elements. Thus $A=C\otimes_{\mathrm{F}}B=B$ is nothing but the algebra $B_{|L|}$.

The para-Bose and para-Fermi number system of Ohnuki and Kamefuchi[7] is also obtained if we put $G=\mathbf{Z}_2\oplus\cdots\oplus\mathbf{Z}_2$ and define σ appropriately.

Returning to the general case, A is a G-graded algebra whose homogeneous component A_α of grade $\alpha\in G$ is the set of elements

$$a=\sum_{\beta+g(M)=\alpha}a_{\beta,M}u_\beta\otimes v_M,$$

where

$$g(M)=\sum_{l\in M}g(l).$$

Here we introduce another gradation on A. An element $a\in A$ given as (2.1) is called *s-homogeneous* of *s-grade M* if $a_{\alpha,M'}=0$, for all $M'\neq M$. Identifying C with $C\otimes 1$ and B with $1\otimes B$, C and B are considered to be subalgebras of A. Here C is nothing but the homogeneous component of A of s-grade $\varnothing$. Moreover, F is a subalgebra of C by identifying $\mathrm{F}u_0$ with F.

Definition 2.1: For $a\in A$ given as (2.1), we define

$$b(a)=\sum_{\alpha\in G_0}a_{\alpha,\varnothing}u_\alpha\otimes 1,\quad s(a)=\sum_{\alpha\in G_0,\,M\neq\varnothing}a_{\alpha,M}u_\alpha\otimes v_M.$$

We call $b(a)$ the *body* of a and $s(a)$ the *soul* of a. Moreover for a subset S of A, we set $b(S)=\{b(a)|a\in S\}$ and $s(S)=\{s(a)|a\in S\}$.

Note that the body $b(a)$ is invertible if it is nonzero and homogeneous, while the soul $s(a)$ is nilpotent. Of course, any element is uniquely decomposed as a sum of its body and its soul. The body of an odd element is zero, but the converse is not true; an even element may have a nonzero soul.

Proposition 2.2: Let A' be a F-submodule of A generated by some elements of the form $u_\alpha\otimes v_M$ with $\alpha\in G_0$ and $M\subset L$. Then the left annihilator and the right annihilator of A' coincide; they are denoted by $\mathrm{Ann}(A')$, where $\mathrm{Ann}(A')$ is a graded submodule of A with respect to s-gradation.

Proof: Let $a=u_\alpha\otimes v_M\in A$ and let $b=\Sigma_{M'\subset L}c_{M'}\otimes v_{M'}$, where $c_{M'}\in C$. Suppose $ba=0$. Then

$$\Big(\sum_{M'}c_{M'}\otimes v_{M'}\Big)(u_\alpha\otimes v_M)=\sum_{M'}\sigma(g(M'),\alpha)c_{M'}u_\alpha\otimes v_{M'}v_M=0.$$

It follows that $c_{M'}=0$, for M' such that $M'\cap M=\varnothing$. This implies $ab=0$ and

$$\mathrm{Ann}(a)=\sum_{M'\cap M\neq\varnothing}C\otimes v_{M'}.$$

Since A' is generated by such a's, our assertion follows.

Definition 2.3: Let A' be as in Proposition 2.2. Define

$$\mathrm{Sav}(A')=\sum_M C\otimes v_M,$$

where M ranges over all the $M\subset L$ such that $v_MA'\neq 0$. Then we have

$$A=\mathrm{Sav}(A')\oplus\mathrm{Ann}(A').$$

For $\alpha_1,\dots,\alpha_n\in G$, $P_{\alpha_1\cdots\alpha_n}$ denotes the projection from A to $\mathrm{Sav}(s(A_{\alpha_1})\cdots s(A_{\alpha_n}))$. Moreover for a G-set $M=\{i_1,\dots,i_n\}$, $P_{g(i_1)\cdots g(i_n)}$ is abbreviated to P_M.

Definition 2.4: Let $I=\{1,\dots,p,p+1,\dots,p+q\}$ be a G-set such that $g(i)$ are even for $i=1,\dots,p$ and odd for $i=p+1,\dots,p+q$. Let $X=A_I=\oplus A_{g(i)}$ be the direct sum of $A_{g(i)}$ $(i\in I)$. The X is a Banach space by the product topology induced from A and is called *superspace* over A.

In the usual supersymmetric case, X is the superspace $B^{p,q}_{|L|}$ of Rogers.[8]

Definition 2.5: Let I and J be G-sets. Let $X=A_I$ and $Y=A_J$ be superspaces. A mapping $T\colon X\to Y$ is an *A-linear mapping*, if there is a $(J\times I)$-matrix $M=(M^j_i)$ such that

$$T(x)^j=\sum_{i\in I}M^j_ix^i,\quad j\in J,$$

for $x=(x^i)\in X$. We say T is associated with M. If M is invertible, then so is T and the inverse mapping T^{-1} is associated with the inverse matrix M^{-1}. Clearly, an A-linear mapping is F-linear. [The definition of a $(J\times I)$-matrix is given in Ref. 4.]

Definition 2.6: For a point $x=(x^i|i\in I)$ in a superspace X, $b(x)=(b(x^i)|i\in I)$ and $s(x)=(s(x^i)|i\in I)$ are called the *body* and the *soul* of x, respectively. Let U be a (connected open) domain of X. Here $b(U)=\{b(x)|x\in U\}$ is called the *body* of U. Then $b(U)$ is contained in the *even part*

$$U_0=\{(x^1,...,x^p,0,...,0)\,|\,(x^1,...,x^p,x^{p+1},...,x^{p+q})\in U\}\text{of } U.$$

Let $x=(x^1,...,x^p,0,...,0)$ be in $b(X)$. Then $x^i=\tilde{x}^i u_{g(i)}$ for some $\tilde{x}^i\in\mathbb{F}$. The mapping $\tilde{}$ which sends x to the point $\tilde{x}=(\tilde{x}^1,...,\tilde{x}^p)$ is a homeomorphism of $b(X)$ onto the (real or complex) p-dimensional space $\mathbb{F}^p$. For a domain V of $b(X)$, $\tilde{V}=\{\tilde{x}|x\in V\}$ is a domain of $\mathbb{F}^p$. We sometimes write the odd coordinates x^{p+j} as ξ^j and express a point of X as $(x,\xi)=(x^1,...,x^p,\xi^1,...,\xi^q)$ in order to distinguish between the even and the odd coordinates.

Let U be a domain of X. Here A^U denotes the set of functions (*superfields*) on U which take their values in A. A function $f\in A^U$ is said to be *homogeneous* of *grade* $\alpha\in G$, if $f(x)\in A_\alpha$, for all $x\in U$. Thus A^U is naturally a σ-commutative G-graded algebra over $\mathbb{F}$.

For a domain V of $b(X)$, A^V also denotes the set of A-valued functions defined on V. A function $f\in A^V$ is written uniquely as

$$f(x)=\sum_{\alpha,M} f_{\alpha,M}(x)u_\alpha\otimes v_M,\quad x\in V, \tag{2.2}$$

where $\alpha\in G_0$, $M\subset L$, and $f_{\alpha,M}(x)\in\mathbb{F}$. The functions $\tilde{f}_{\alpha,M}$, which are defined by

$$\tilde{f}_{\alpha,M}(\tilde{x})=f_{\alpha,M}(x),\quad x\in V,$$

are $\mathbb{F}$-valued functions on the domain $\tilde{V}$.

III. DIFFERENTIAL CALCULUS

In this section we study differential calculus on the superspaces in the sense of Sec. II. Through this and the subsequent sections we assume that the base field $\mathbb{F}$ is the real number field $\mathbb{R}$.

Let $I=\{1,...,p+q\}$ be a G-set with p even and q odd elements, and let $X=A_I$ be the superspace. Let U be a (connected open) domain of X.

Definition 3.1: (Compare Rogers.[8]) Let $i\in I$ and a point $x_0\in U$ be fixed. A function $f\in A^U$ is called *right differentiable* at x_0 with respect to x^i, if there is a constant $a\in A$ such that

$$f(x_0^1,...,x_0^{i-1},x_0^i+y,x_0^{i+1},...,x_0^{p+q})$$
$$=f(x_0)+ay+o(\|y\|)\quad(\text{as } y\to 0), \tag{3.1}$$

for

$$y\in A_{g(i)}$$

with

$$(x_0^1,...,x_0^{i-1},x_0^i+y,x_0^{i+1},...,x_0^{p+q})\in U.$$

The constant a is called the *right differential coefficient* of f at x_0 with respect to x^i.

The function f is right differentiable on U with respect to x^i if there is a function f' in A^U such that $f'(x)$ is a right differential coefficient of f at any x in U. Though f', which is called a *right derivative* of f, is not uniquely determined in general, it is unique modulo $\mathrm{Ann}(A_{g(i)})$. We use the symbol f_{x^i} or simply $f_{(i)}$ to denote one of them.

Remark 3.2: If $i\in I$ is an even index and if $f\in A^U$ is right differentiable at x_0 with respect to x^i, then the right-differential coefficient is (absolutely) unique and is equal to

$$\lim_{\delta\to 0}\frac{1}{\delta}\,[\,f(x_0^1,...,x_0^{i-1},x_0^i+\delta u_{g(i)},x_0^{i+1},...,x_0^n)$$
$$-f(x_0)\,]u_{g(i)}{}^{-1}, \tag{3.2}$$

where $\delta\in\mathbb{R}$.

Definition 3.3: A function $f\in A^U$ is r-times right differentiable on U with respect to $x^{i_1},...,x^{i_r}$, if there is an $(r-1)$th derivative $f_{(i_1\cdots i_{r-1})}$ with respect to $x^{i_1},...,x^{i_{r-1}}$, which is right differentiable on U with respect to x^{i_r}; a derivative of $f_{(i_1,...,i_{r-1})}$ with respect to x^{i_r} is denoted by $f_{(i_1,...,i_r)}$.

Proposition 3.4: The difference of two derivatives of f with respect to $x^{i_1},...,x^{i_r}$ belongs to $\mathrm{Ann}(\Pi_{j=1}^r A_{g(i_j)})$.

Proof: Let h and g be derivatives of f with respect to x^i, then $k=h-g\in\mathrm{Ann}(A_{g(i)})$. Hence

$$\lim_{\delta\to 0}\frac{1}{\delta}\,[k(x^1,...,x^{j-1},x^j+\delta y^j,x^{j+1},...,x^{p+q})-k(x)]$$
$$=k_{(j)}(x)y^j$$

is also in $\mathrm{Ann}(A_{g(i)})$, where $\delta\in\mathbb{R}$. Because y^j is arbitrary in $A_{g(j)}$, $k_{(j)}(x)$ belongs to $\mathrm{Ann}(A_{g(i)}A_{g(j)})$. Since $h_{(j)}-g_{(j)}$ is equal to $k_{(j)}$ modulo $\mathrm{Ann}(A_{g(j)})$, $h_{(j)}-g_{(j)}$ belongs to $\mathrm{Ann}(A_{g(i)}A_{g(j)})$.

Repeating this argument we can show the assertion for higher-order derivatives.

By definition it is clear that an $(r-s)$th derivative $(f_{(i_1,...,i_s)})_{(i_{s+1},...,i_r)}$ of an sth derivative $f_{(i_1,...,i_s)}$ of f is an rth derivative $f_{(i_1,...,i_r)}$ of f. Therefore

$$f_{(i_1,...,i_r)}$$
$$=(f_{(i_1,...,i_s)})_{(i_{s+1},...,i_r)}\quad\mathrm{mod}\ \mathrm{Ann}\Big(\prod_{j=1}^r A_{g(i_j)}\Big).$$

Definition 3.5: A continuous function f in A^U is called G^0on U The function f is said to be G^r on U if there is a continuous rth right derivative $f_{(i_1,...,i_r)}$ on U for any $i_1,...,i_r\in I$.

The set of G^r-functions on U is denoted by $G^r(U)$. It is a σ-commutative algebra over $\mathbb{R}$ in a natural way (cf. Proposition 3.10). In a usual way we can prove the following proposition.

Proposition 3.6: Let f be in A^U. Then f is G^r on U if and only if

$$f(x+y)=f(x)+\sum_{i=1}^{p+q}h_i(x)y^i+o(\|y\|)\quad(\text{as } y\to 0),$$

for some G^{r-1}- functions h_i, where x, $x+y\in U$. This $h_i(x)$ is a right derivative of f with respect to x^i.

Let J be a G-set and $f=\{f^j|\,j\in J\}$ be a set of functions $f^j\in G^r(U)$ such that the grade of f^j is $g(j)$. Then f induces a G^r-mapping $f=(f^j)$: $U\to Y=A_J$ by

$$f(x)=(f^j(x)),\quad x\in U.$$

Preposition 3.7: Let $y=(y^j)$ be a G^1-mapping from U to a domain V in A_J. Then for $f\in G^1(V)$, the composition $f(y(x))$ of f and y belongs to $G^1(U)$ and the equality

$$f_{(i)}(x) = \sum_j f_{(j)}(y(x)) \cdot y^j_{(i)}(x) \quad \text{mod Ann}(A_{g(i)}) \tag{3.3}$$

holds on U.

Proof: By definition

$$y^j(x^1,...,x^{i-1},x^i + x^i_1,x^{i+1},...,x^{p+q}) = y^j(x) + y^j_{(i)}(x) \cdot x^i_1 + o(\|x^i_1\|),$$

and by Proposition 3.6

$$f(y + y_1) = f(y) + \sum_{j\in J} f_{(j)}(y) \cdot y^j_1 + o(\|y_1\|).$$

Hence

$$f(y(x^1,...,x^{i-1},x^i + x^i_1,x^{i+1},...,x^{p+q})) = f(y(x)) + \sum_j f_{(j)}(y(x)) \cdot y^j_{(i)}(x) \cdot x^i_1 + o(\|x^i_1\|).$$

This completes the proof.

Proposition 3.8 (Taylor's expansion): Let $y = (y^i) \in s(X)$ be the nilpotent of nilpotency r, that is, $y^{i_1} \cdots y^{i_r} = 0$ for any r-tuple $(i_1,...,i_r)$, $i_k \in I$. Let $f \in G^r(U)$ and $x \in U$. Suppose $x + \theta y \in U$ for all θ with $0 < \theta \leqslant 1$, then

$$f(x+y) = \sum_{n=0}^{r-1} \frac{1}{n!} \sum_{(i_1,...,i_n)} f_{(i_n \cdots i_1)}(x) y^{i_1} \cdots y^{i_n}. \tag{3.4}$$

Proof: Let $a \in \mathbb{R}$ and set $g(a) = f(x + ay)$. Then $g(a)$ is a function defined on a neighborhood of 0 in $\mathbb{R}$ and takes its values in A. By the ordinary Taylor theorem,

$$g(1) = \sum_{n=0}^{r-1} \frac{1}{n!} g^{(n)}(0) + \frac{1}{r!} g^{(r)}(\theta),$$

where $0 < \theta < 1$. Applying (3.3) repeatedly [note that (3.3) is an exact equality when i is even], we have

$$g^{(n)}(0) = \sum_{(i_1,...,i_n)} f_{(i_n \cdots i_1)}(x) \cdot y^{i_1} \cdots y^{i_n}.$$

Since $g^{(r)}(\theta) = 0$, (3.4) follows.

Proposition 3.9: For $f \in G^2(U)$, we have

$$f_{(ij)} = \sigma(g(i),g(j)) f_{(ji)} \quad \text{mod Ann}(A_{g(i)} A_{g(j)}).$$

Proof: Set

$$g(a,b) = f(x^1,...,x^{i-1},x^i + ay^i,x^{i+1}, ...,x^{j-1},x^j + by^j,x^{j+1},...,x^{p+q}),$$

where, a, $b \in \mathbb{R}$. Then the equations $g_{a,b}(0,0) = f_{(ij)}(x) y^j y^i$ and $g_{b,a}(0,0) = f_{(ji)}(x) y^i y^j$ hold by Proposition 3.7. Since $g_{a,b}(0,0) = g_{b,a}(0,0)$, $f_{(ij)}(x) - \sigma(g(i), g(j)) f_{(ji)}(x)$ lies in $\text{Ann}(A_{g(i)} A_{g(j)})$.

Proposition 3.10: Let f, $g \in G^1(U)$ and h be homogeneous of grade α. Then $fg \in G^1(U)$ and we have

$$(fh)_{(i)} = \sigma(g(i),\alpha) f_{(i)} \cdot h + f \cdot h_{(i)} \quad \text{mod Ann}(A_{g(i)}).$$

Proof: Straightforward from the definition.

Proposition 3.11: Let f be a G^r-function on U let a be a point in X that is nilpotent of nilpotency r. Then $f_{(i_n \cdots i_1)}(x) \cdot a^{i_1} \cdots a^{i_n}$ is a G^r-function on U.

Proof: By Propositions 3.8 and 3.9 we have

$$f(x + ta) = \sum_n \sum_{i_1 < \cdots < i_n} c_{i_1 \cdots i_n} f_{(i_n \cdots i_1)}(x) \cdot a^{i_1} \cdots a^{i_n} t^{i_1} \cdots t^{i_n},$$

for any sufficiently small t in $\mathbb{R}^{p+q}$, where $c_{i_1 \cdots i_n}$ are nonzero real constants. Therefore every $f_{(i_n \cdots i_1)}(x) \cdot a^{i_1} \cdots a^{i_n}$ can be written as a linear combination over $\mathbb{R}$ of G^r-functions $f_t(x) = f(x + ta)$ with $t \in \mathbb{R}^{p+q}$. Consequently $f_{(i_n \cdots i_1)}(x) \cdot a^{i_1} \cdots a^{i_n}$ is G^r on U.

Definition 3.12: For a subset S of X, we define $\overline{S} = \{x + a | x \in S, a \in s(X)\}$ and call it the *(soul) saturation*[11] of S. If $S = \overline{S}$, then S is called *saturated*; S is called *s-connected*[12] if $S \cap \overline{\{x\}}$ is connected for every $x \in S$.

Proposition 3.13: Suppose a domain U is s-connected and every point of $s(U)$ has nilpotency r. Then any G^r-function f on U can be uniquely extended to a G^r-function $\bar{f}$ on $\overline{U}$.

Proof: For $y = x + a \in \overline{U}$ with $x \in U$ and $a \in s(X)$, define

$$\bar{f}(y) = \bar{f}(x + a) = \sum_{n=0}^{r-1} \frac{1}{n!} \sum_{(i_1,...,i_n)} f_{(i_n \cdots i_1)}(x) a^{i_1} \cdots a^{i_n}.$$

The s-connectedness of U assures that $\bar{f}$ is well defined. We see $\bar{f}|_U = f$ and $\bar{f} \in G^r(\overline{U})$ by Proposition 3.11, and moreover $\bar{f}$ is unique because of Propositions 3.4 and 3.8.

As the following example shows, the assumption that U is s-connected cannot be removed in Proposition 3.13. The example also explains the necessity of the condition that $x + \theta y \in U$ for all θ with $0 < \theta \leqslant 1$ in Proposition 3.8.

Example 3.14: Let A be the Grassmann algebra generated by two elements v_1, v_2 and $X = \mathbb{R} + \mathbb{R} v_1 v_2$ be the superspace over A with one even coordinate. Let

$$U = \{x + y v_1 v_2 | x, y \in \mathbb{R} \text{ and } y \neq 0, \text{ for } x \leqslant 0\},$$

then U is a domain of X which is not s-connected. Let $\phi(x)$ be a C^∞-function on $\mathbb{R}$ such that $\phi(x) = 1$, for $x \geqslant 0$ and $\phi(x) = 0$, for $x \leqslant -1$. Define a function f on U by

$$f(x + y v_1 v_2) = \begin{cases} 1, & \text{if } x \leqslant 0 \text{ and } y > 0, \\ \phi(x) + \left(\dfrac{d\phi(x)}{dx}\right) y v_1 v_2, & \text{otherwise.} \end{cases}$$

Then f is a G^∞-function on U. However, since $f = 0$ on $\{x + y v_1 v_2 | x < -1, y < 0\}$ and $f = 1$ on $\{x + y v_1 v_2 | x < -1, y > 0\}$, f cannot be extended to $\overline{U} = X$ as a continuous function, still less as a G^∞-function.

Left differentiability and *left derivatives* are defined dually. It is clear that a function f is left differentiable if and only if it is right differentiable.

IV. FUNCTIONS ON THE BODY

In this section we study functions defined on the body. Let $X = A_I$ be the superspace and V be a domain of the body $b(X)$ of X. Since we treat only even variables here, we suppose the G-set I is even, that is, $I = I_0 = \{1,..., p\}$.

Let f be an A-valued function defined on V. The right differentiability and the right differential coefficient of f are defined by Eq. (3.1) in Definition 3.1 under the restriction that $q = 0$, $y \in C_{g(i)}$, and $(x^1_0,...,x^{i-1}_0,x^i_0 + y,x^{i+1}_0,...,x^{p+q}_0)$

$\in V$. Since the right differential coefficient of f at x_0 with respect to x^i is unique, as mentioned in Remark 3.2, we use the notation $\partial f(\partial x^i)^{-1}(x_0)$ to denote it.

As we stated at the end of Sec. II, $f\in A^V$ is a sum of $f_{\alpha,M}$, $\alpha \in G_0$, $M \subset L$, with which we associate $\mathbb{R}$-valued functions $\tilde{f}_{\alpha,M}$ on $\tilde{V}\subset \mathbb{R}^p$.

Proposition 4.1: $f\in A^V$ is right differentiable at x with respect to x^i if and only if $\tilde{f}_{\alpha,M}$ is differentiable at $\tilde{x}$ with respect to $\tilde{x}^i$ for every $\alpha\in G_0$ and $M\subset L$, and

$$\partial f(\partial x^i)^{-1}(x) = \sum_{\alpha\in G, M\subset L} \frac{\partial \tilde{f}_{\alpha,M}}{\partial \tilde{x}^i}(\tilde{x})(u_\alpha \otimes v_M)u_{g(i)}^{-1}.$$

For a positive integer n, $[\partial(\partial x^i)^{-1}]^n$ means n times application of right differentiation with respect to x^i. Let $N=\{n_i\}_{i\in I}$ be a sequence of non-negative integers indexed by I. Set $|N| = \Sigma_{i\in I} n_i$ and $N! = \Pi_{i\in I}(n_i!)$. We use the abbreviation x^N for the ordered product $\Pi(x^i)^{n_i}$. The higher-order derivative $\partial^{|N|} f(\partial x^N)^{-1}$ of f is a successive application of $[\partial(\partial x^p)^{-1}]^{n_p}$, $[\partial(\partial x^{p-1})^{-1}]^{n_{p-1}},\ldots,[\partial(\partial x^1)^{-1}]^{n_1}$ to f in this order. Here $f\in A^V$ is said to be C^r if it is r-times continuously right differentiable; $C^r(V)$ stands for the set of C^r-functions on V.

By Propositions 4.1 we have the following proposition.

Proposition 4.2: $f\in A^V$ is C^r if and only if every $\tilde{f}_{\alpha,M}$ is C^r.

Definition 4.3: Let J be a G-set and let V be a domain in $b(X)$. Let $f=(f^j)$ be a C^1-mapping from V to A_J. The $(J\times I)$-matrix

$$D(f/x) = (\partial f^j(\partial x^i)^{-1})$$

is called the Jacobian matrix for f; if J is even and $|I| = |J|$, the determinant of $D(f/x)$ is called the Jacobian for f, and is denoted by $\Delta(f/x)$.

Hereafter we consider the case where J is even, $|I| = |J|$ and each f^j is a C-valued C^1-function, that is, $f^j(x)$ $\tilde{f}^j(\tilde{x})u_{g(j)}$. Then f is a mapping from V to the body $b(Y)$ of $Y=A_J$. By Proposition 4.1 we have

$$\partial f^j(\partial x^i)^{-1} = \frac{\partial \tilde{f}^j}{\partial \tilde{x}^i} u_{g(j)} u_{g(i)}^{-1}. \tag{4.1}$$

Let $\tilde{I}$ (resp. $\tilde{J}$) be the G-set obtained from I (resp. J) by redefining the grade of every element of I (resp. J) to be 0. Let U_I (resp. U_J) be the $(\tilde{I}\times I)$- [resp. $(J\times\tilde{J})$-] matrix defined by $(U_I)^i_{i'} = \delta^i_{i'} u_{g(i)}$ [resp. $(U_J)^j_{j'} = \delta^j_{j'} u_{g(j)}$].

Let $D(\tilde{f}/\tilde{x})$ and $\Delta(\tilde{f}/\tilde{x})$ be the ordinary Jacobian matrix and Jacobian for $\tilde{f}$, respectively. Then we have by (4.1)

$$D(f/x) = U_J D(\tilde{f}/\tilde{x})U_I^{-1}. \tag{4.2}$$

Thus by Theorem 3.6 of Ref. 4 we have

$$\Delta(f/x) = \det(U_J)\Delta(\tilde{f}/\tilde{x})\det(U_I)^{-1}$$
$$= \Delta(\tilde{f}/\tilde{x})\det(U_J)\det(U_I)^{-1}.$$

This formula implies the following proposition.

Proposition 4.4: Under the same situation as above, $D(f/x)$ is invertible if and only if $D(\tilde{f}/\tilde{x})$ is invertible. Moreover if $I=J$, then we have

$$\Delta(f/x) = \Delta(\tilde{f}/\tilde{x}).$$

Next we define integration on the body.

Definition 4.5: Let W be the subset of V such that $\tilde{W}$ is a measurable subset of $\mathbb{R}^p$. Let $f\in A^V$ be integrable on W, that is, all the $\tilde{f}_{\alpha,M}$ are integrable on $\tilde{W}$. We define

$$\int_W f(x)dx = \int_W f(x^1,\ldots,x^p)dx^1\cdots dx^p$$
$$= \sum_{\alpha,M}\left[\int_{\tilde{W}} \tilde{f}_{\alpha,M}(\tilde{x}^1,\ldots,\tilde{x}^p)d\tilde{x}^1\cdots d\tilde{x}^p\right](u_\alpha\otimes v_M).$$

The following proposition is a special case of Theorem 7.4 and will be used to prove it in Sec. VII.

Proposition 4.6: Let I be an even G-set and V and W be domains in $b(X)$, $X=A_I$. Let $y=(y^i)$ be a one to one C^1-mapping from W onto V such that $\Delta(y/x)$ is nonzero. Then we have

$$\int_W f(y(x))\left|\Delta\left(\frac{y}{x}\right)\right|dx = \int_V f(y)dy.$$

Proof: By definition 4.5 and Proposition 4.4 we have

$$\int_W f(y(x))\left|\Delta\left(\frac{y}{x}\right)\right|dx$$
$$= \sum_{\alpha,M}\left[\int_{\tilde{W}} \tilde{f}_{\alpha,M}(\tilde{y}(\tilde{x}))\left|\Delta\left(\frac{\tilde{y}}{\tilde{x}}\right)\right|d\tilde{x}\right](u_\alpha\otimes v_M). \tag{4.3}$$

By the usual formula of change of coordinates, (4.3) is equal to

$$\sum_{\alpha,M}\left[\int_{\tilde{V}}\tilde{f}_{\alpha,M}(\tilde{y})d\tilde{y}\right](u_\alpha\otimes v_M) = \int_V f(y)dy.$$

The following propositions hold.

Proposition 4.7: If $f\in C^1(V)$ has a compact support then we have

$$\int_V\{\partial f(\partial x^i)^{-1}(x)\}dx = 0.$$

Proposition 4.8: Let f, $h\in C^1(V)$ and suppose h is homogeneous of grade β. If h has a compact support, then we have

$$\int_V\{f(x)\partial h(\partial x^i)^{-1}(x)\}dx$$
$$= -\sigma(g(i),\beta)\int_V\{\partial f(\partial x^i)^{-1}(x)h(x)\}dx.$$

V. JACOBIAN AND INVERSE MAPPING THEOREM

In this section G-sets are not necessarily even. The even (resp. odd) part of a G-set I is denoted by I_0 (resp. I_1).

Definition 5.1: Let I and J be G-sets. Let $X=A_I$, $Y=A_J$, and U be a domain of X. Let $f=(f^j|j\in J)$ be a G^1-mapping from U to Y. The $(J\times I)$-matrix

$$D(f/x) = (f^j_{(i)}(x))$$

is called the *Jacobian matrix* for f, where $f^j_{(i)}$ is a right derivative of f defined in Definition 3.1. Note that it is unique up to modulo $\mathrm{Ann}(A_{g(i)})$. When $D(f/x)$ is square, that is, $|I_0| = |J_0|$ and $|I_1| = |J_1|$, the superdeterminant of $D(f/x)$ is called the *Jacobian* for f and denoted by $\Delta(f/x)$. It is unique up to modulo $\mathrm{Ann}(s(A))$. Hence the invertibility of $D(f/x)$ does not depend on the choice of the derivatives $f^j_{(i)}$. (See Definition 3.9 and Proposition 3.10 of Ref. 4.)

Proposition 5.2: Let I, J, and K be G-sets and let $U\subset A_I$ and $V\subset A_J$. Let $f=(f^k)$ be a G^1-mapping from V to A_K and $y=(y^j)$ be a G^1-mapping from U to V. Composing f and

y, f is considered to be a mapping from U to A_K. Then

$$D(f/x) \simeq D(f/y) \cdot D(y/x),$$

where $\simeq$ means that the (k,i)-elements of the matrices are equal modulo $\mathrm{Ann}(A_{g(i)})$. Thus if the matrices are square,

$$\Delta(f/x) = \Delta(f/y) \cdot \Delta(y/x) \quad \mathrm{mod}(s(A)).$$

Proof: It immediately follows from Proposition 3.7 and Theorem 3.11 of Ref. 4.

Corollary 5.3: In the same situation as above, $D(f/x)$ is invertible if and only if both $D(f/y)$ and $D(y/x)$ are invertible.

Theorem 5.4 (inverse mapping theorem): Let I and J be G-sets such that $|I_0| = |J_0|$, $|I_1| = |J_1|$. Let $X = A_I$, $Y = A_J$, and $U \subset X$. Let $f = (f^j)$ be a G^r-mapping from U to Y with $r \geqslant 1$. If $\Delta(f/x) \neq 0$ at $x_0 \in U$, then there are an open neighborhood V of $f(x_0)$, a neighborhood W of x_0, and a G^r-mapping $h: V \to W$ such that $h(f(x)) = x$ and $f(h(y)) = y$, for all $x \in W$ and $y \in V$.

Proof: With the Jacobian matrix $D = D(f/x)$ we associate an invertible A-linear mapping T from X to Y (Definition 2.5). Let $T_{\mathbb{R}}$ denote the $\mathbb{R}$-linear mapping T regarding X and Y as vector spaces over $\mathbb{R}$. Regarding f as a mapping from the domain U in the $\mathbb{R}$-vector space X to Y, its ordinary Jacobian matrix corresponds to $T_{\mathbb{R}}$. Since $(T_{\mathbb{R}})^{-1} = (T^{-1})_{\mathbb{R}}$, the usual inverse mapping theorem gives us a neighborhood V of $f(x_0)$, a neighborhood W of x_0 and a mapping h from V to W such that $h(f(x)) = x$ and $f(h(y)) = y$, for $x \in W$ and $y \in V$. Since the ordinary Jacobian matrix of h corresponds to $(T_{\mathbb{R}})^{-1} (= (T^{-1})_{\mathbb{R}})$ and T^{-1} is associated with the inverse D^{-1} of D, we have

$$h(y+z) - h(y) = T^{-1}(x) + o(\|z\|) = D^{-1}z + o(\|z\|).$$

It follows that h is a G^1-mapping from V to U of which (super) Jacobian matrix is D^{-1}. Moreover, all the elements of D^{-1} are rational functions of the derivatives of f and the denominators of the rational functions have nonzero bodies (see Corollary 3.8 of Ref. 4). They are clearly G^{r-1}-functions. Consequently, h is a G^r-function as desired.

Proposition 5.5: Under the same condition as above, suppose $\Delta(f/x) \neq 0$ and U and every point of $s(U)$ has nilpotency r.

(1) If U is saturated, then so is $f(U)$.

(2) If $f(x)$ is one to one on U, then so is $b(f(x))$ on $b(U)$.

Proof: By the inverse mapping theorem, for any $y_0 \in f(U)$ there are an s-connected open neighborhood V of y_0 in Y and a G^r-mapping $h: V \to U$ such that $f(h(y)) = y$, for all $y \in V$. By Proposition 3.13, h can be extended to the saturation $\overline{V}$, and $f(h(y)) = y$ holds on $\overline{V}$ because $f \cdot h$ is G^r. Therefore $\overline{V} = f(h(\overline{V})) \subset f(U)$ and (1) follows.

Assume that $b(f(x_1)) = b(f(x_2))$ for some x_1 and x_2 in $b(U)$. Then there are s-connected neighborhoods V_i of $y_i = f(x_i)$ and G^r-mappings h_i on V_i to U such that $h_i(y_i) = x_i$ and $f(h_i(y)) = y$, for $y \in V_i$ $(i = 1,2)$. Thus h_i can be extended to $\overline{V}_1 \cap \overline{V}_2$, both of which are the inverse of f. Since f is one to one, $h_1 = h_2 (= h)$ holds. Since $b(y_1) = b(y_2)$ and h is G^r, we have $b(h(y_1)) = b(h(y_2))$. It follows that $x_1 = x_2$ and (2) is proved.

VI. STANDARD EXPANSIONS

Let I be a G-set and $X = A_I$. Let U be a s-connected domain of X. Let r be a fixed integer such that the soul of every point of X has nilpotency r. Since every G^r-function on U is uniquely extended to the saturation $\overline{U}$ of U, we assume $U = \overline{U}$ in this section.

Definition 6.1: Let f be a function defined on $b(U)$ and t be a non-negative integer. This f is called *excessively* C^t or $C^t_!$ if for every $i_1,...,i_n \in I_0$ with $0 \leqslant n \leqslant r$, $P_{g(i_1)\cdots g(i_n)} f$ is n-times right differentiable with respect to $x^{i_1},...,x^{i_n}$ and the derivative $\partial^n(P_{g(i_1)\cdots g(i_n)} f)(\partial x^{i_1})^{-1} \cdots (\partial x^{i_n})^{-1}$ is C^t (recall Definition 2.3).

When $n = 0$, $P_{g(i_1)\cdots g(i_n)}$ in the above definition is the identity mapping and so a $C^t_!$-function is C^t.

Proposition 6.2: If f is $C^t_!$ on $b(U)$ with $t \geqslant 1$, then $\partial f(\partial x^{i_0})^{-1}$ is $C^{t-1}_!$ on $b(U)$ for $i_0 \in I_0$.

Proof: If f is $C^t_!$,

$$\partial^n[P_{g(i_1)\cdots g(i_n)} f](\partial x^{i_1})^{-1} \cdots (\partial x^{i_n})^{-1}$$

is C^t for every $i_1,...,i_n \in I_0$ with $n \leqslant r$. Hence

$$\partial^n[P_{g(i_1)\cdots g(i_n)}(\partial f(\partial x^{i_0})^{-1})](\partial x^{i_1})^{-1} \cdots (\partial x^{i_n})^{-1}$$
$$= \epsilon\partial\left[\partial^n(P_{g(i_1)\cdots g(i_n)} f)(\partial x^{i_1})^{-1} \cdots (\partial x^{i_n})^{-1}\right](\partial x^{i_0})^{-1}$$

is C^{t-1} for $i_1,...,i_n \in I_0$ with $n \leqslant r$, where

$$\epsilon = \Pi_{k=1}^n \sigma(g(i_0), g(i_k)).$$

This implies $\partial f(\partial x^{i_0})^{-1}$ is $C^{t-1}_!$.

Proposition 6.3: If f is a G^r-function defined on the even part U_0 of U, then the restriction $f|_{b(U)}$ of f to $b(U)$ is $C^r_!$. Conversely, any $C^r_!$ function on $b(U)$ can be extended uniquely to a G^r-function on U_0.

Proof: Let $f \in G^r(U_0)$ and $i_1,...,i_n \in I_0$ be given. Set $g(x) = f_{(i_1\cdots i_n)}(x)$. By Proposition 3.11, $g(x) \cdot a^{i_n} \cdots a^{i_1}$ is G^r on U_0, for any $a = (a^{i_j})$, $a^{i_j} \in s(A_{g(i_j)})$. Write

$$g(x) = \sum_{\alpha,M} g_{\alpha,M}(x)(u_\alpha \otimes v_M), \quad \text{for } x \in b(U),$$

where $g_{\alpha,M}(x) \in \mathbb{R}$. Then

$$P_{g(i_1)\cdots g(i_n)}(g(x)) = \sum_{\alpha,M'} g_{\alpha,M'}(x)(u_\alpha \otimes v_{M'}),$$

where M' ranges over all $M' \subset L$ such that $v_{M'} \cdot s(A_{g(i_1)}) \cdots s(A_{g(i_n)}) \neq 0$. For any such M' take $a_0 = (a_0^{i_j})$ so that $v_{M'} \cdot a_0^{i_1} \cdots a_0^{i_n} \neq 0$. Let $\{M''\}$ be the family of subsets M'' of L satisfying $v_{M''} \cdot a_0^{i_1} \cdots a_0^{i_n} \neq 0$. Since

$$g(x) \cdot a_0^{i_n} \cdots a_0^{i_1} = \sum_{\alpha,M'} g_{\alpha,M'}(x)(u_\alpha \otimes v_{M'})a_0^{i_n} \cdots a_0^{i_1}$$
$$= \sum_{\alpha,M''} g_{\alpha,M''}(x)(u_\alpha \otimes v_{M''})a_0^{i_n} \cdots a_0^{i_1}$$

is C^r on $b(U)$, every $g_{\alpha,M''}$ is C^r by Proposition 4.1, and in particular $g_{\alpha,M'}(x)$ is C^r. Therefore

$$P_{g(i_1)\cdots g(i_n)}(g(x)) = \partial^n[P_{g(i_1)\cdots g(i_n)} f](\partial x^{i_1})^{-1} \cdots (\partial x^{i_n})^{-1}$$

is C^r and f is $C^r_!$ on $b(U)$.

To show the converse, we shall prove more generally that if h is a $C_{\uparrow}^{t}$-function on $b(U)$, then the function f on U_0 defined by

$$f(x)=\sum\frac{1}{N!}\,[\partial^{|N|}(P_Nh)(\partial x^N)^{-1}(b(x))]s(x)^N,\quad x\in U_0,$$

is G^t, where $N=\{n_i\}$ is a sequence of non-negative integers indexed by I_0 and P_N is the projection to $\mathrm{Sav}[\Pi_{i\in I_0}(s(A_{g(i)}))^{n_i}]$. We proceed by induction on t. When $t=0$, every $[\partial^{|N|}(P_Nh)(\partial x^N)^{-1}(b(x))]s(x)^N$ is continuous and so is $f(x)$. Suppose $t>0$. Let

$$f_i(x)=\sum\frac{1}{N!}\,[\partial^{|N|}\{P_N(\partial h(\partial x^i)^{-1})\}\times(\partial x^N)^{-1}(b(x))]s(x)^N,$$

for $x\in U_0$. Then we can see

$$f(x^1,\ldots,x^{i-1},x^i+y^i,x^{i+1},\ldots,x^p)-f(x)=f_i(x)y^i+o(\|y^i\|),$$

and we find that $f(x)$ is right differentiable with respect to x^i on U_0 and its derivative $f_{(i)}(x)$ is equal to $f_i(x)$. Because every $\partial h(\partial x^i)^{-1}$ is $C_{\uparrow}^{t-1}$ by Proposition 6.2, f_i is G^{t-1} by induction hypothesis. Consequently we find f is G^t.

The uniqueness of the extension follows from Proposition 3.8.

Example 6.4: Let X be the superspace given in Example 3.14. Then the soul of every point of X has nilpotency 2. Let

$$g(x)=g_0(x)+g_1(x)v_1+g_2(x)v_2+g_3(x)v_1v_2$$

be a function on $b(X)=\mathbb{R}$, where $g_i(x)$ are real-valued functions. Then g is $C_{\uparrow}^2$ if and only if g_0 is C^3 and g_1, g_2, and g_3 are C^2. In this case if we define

$$f(x)=g(b(x))+g'(b(x))s(x),\quad\text{for } x\in X,$$

then f is a G^2-function on X by Proposition 6.3.

Now let f be a G^r-function defined on the whole domain U. By Proposition 3.8 we have

$$f(x,\xi)=\sum\frac{1}{m!n!}\sum_{(i_1,\ldots,i_n;j_1,\ldots,j_m)}f_{\xi^{j_m}\cdots\xi^{j_1}x^{i_n}\cdots x^{i_1}}(b(x))\times s(x^{i_1})\cdots s(x^{i_n})\xi^{j_1}\cdots\xi^{j_m},$$

for $x\in U_0$, $\xi\in U_1$. By the proof of Proposition 6.3, $P_{g(j_1)\cdots g(j_m)g(i_1)\cdots g(i_n)}(f_{\xi^{j_m}\cdots\xi^{j_1}x^{i_n}\cdots x^{i_1}})$ is C^r, and hence $P_{g(j_1)\cdots g(j_m)}(f_{\xi^{j_m}\cdots\xi^{j_1}})$ is $C_{\uparrow}^r$ on $b(U)$. It is extended uniquely to the G^r-function $f_{j_1\cdots j_m}$ on U_0, which is actually given by

$$f_{j_1\cdots j_m}(x)=\frac{1}{m!}\sum_n\frac{1}{n!}\sum_{(i_1,\ldots,i_n)}P_{g(j_1)\cdots g(j_m)}[f_{\xi^{j_m}\cdots\xi^{j_1}x^{i_n}\cdots x^{i_1}}(b(x))]\times s(x^{i_1})\cdots s(x^{i_n}),$$

for $x\in U_0$. Therefore we have

$$f(x,\xi)=\sum f_{j_1\cdots j_m}(x)\xi^{j_1}\cdots\xi^{j_m}.$$

Summarizing the previous argument, we have the following theorem.

Theorem 6.5: Let f be a G^r-function on U, then it can be uniquely expressed as follows:

$$f(x,\xi)=\sum_M f_M(x)\xi^M,\quad(x,\xi)\in U,\tag{6.1}$$

where M ranges over all subsets of I_1 and

(i) $f_M(x)$ is a G^r-function on U_0,

(ii) $f_M(x)$ belongs to $\mathrm{Sav}(\Pi_{j\in M}A_{g(j)})$, for all $x\in b(U)$.

As we used in (6.1), ξ^M means $\xi^{j_1}\cdots\xi^{j_m}$, for $M=\{j_1,\ldots,j_m\}\subset I_1$ with $j_1<\cdots<j_m$.

Definition 6.6: The expression (6.1) in Theorem 6.5 is called the *standard expansion*[13] of $f(x,\xi)$. The function $f_{I_1}(x)$ that appeared in (6.1) is called the *top* of $f(x,\xi)$. If every $f_M(x)$ in (6.1) belongs to the body $b(A)$ for $x\in b(U)$, $f(x,\xi)$ is called *proper*.

As we stated before, derivatives with respect to odd variables are not unique, but here we can choose canonical ones. Using the standard expansion of f, we have

$$\begin{aligned}f(x,\xi^1,\ldots,\xi^{j-1},\xi^j+\eta^j,\xi^{j+1},\ldots,\xi^q),&=\sum_{j\notin M}f_M(x)\xi^M+\sum_{j\in M}\epsilon_Mf_M(x)\xi^{M-\{j\}}(\xi^j+\eta^j)\\&=f(x,\xi)+\sum_{j\in M}\epsilon_Mf_M(x)\xi^{M-\{j\}}\eta^j,\end{aligned}$$

where $\epsilon_M=\Pi_{j<k}\sigma(g(j),g(k))$. This shows that $\Sigma_{j\in M}\epsilon_Mf_M(x)\xi^{M-\{j\}}$ is a derivative of f with respect to ξ^j. We call it *the* derivative of f with respect to ξ^j, which is denoted by $\partial f(\partial\xi^j)^{-1}$, that is,

$$\partial f(\partial\xi^j)^{-1}=\sum_{M\subset I_1,j\in M}\epsilon_Mf_M(x)\xi^{M-\{j\}}.\tag{6.2}$$

Since the derivative of f with respect to even variable x^i is unique, we also use the notation $\partial f(\partial x^i)^{-1}$ for it. It is clear that

$$\partial f(\partial x^i)^{-1}=\sum_M\sigma(g(i),g(M))\partial f_M(\partial x^i)^{-1}(x)\xi^M.\tag{6.3}$$

Since $\partial f(\partial\xi^j)^{-1}$ is again a G^r-function and (6.2) is its standard expansion, the derivative of it can be defined. On the other hand $\partial f(\partial x^i)^{-1}$ may not be G^r, but considering (6.3) as if it is the standard expansion of $\partial f(\partial x^i)^{-1}$, we define its derivative in the same way as above. Inductively we can define the higher-order derivative $\partial^Nf(\partial x^N)^{-1}$ of f for a sequence $N=\{n_i\}$ of non-negative integers indexed by I.

If $f(x,\xi)=\Sigma_Mf_M(x)\xi^M$ and $h(x,\xi)=\Sigma_Mh_M(x)\xi^M$ are the standard expansions of f and h, then $\Sigma_M(f_M(x)+h_M(x))\xi^M$ is the standard expansion of $f+h$. From this fact it follows that the operation $\partial(\partial x^i)^{-1}$ is additive.

The left derivative $(\partial x^i)^{-1}f\partial$ of $f\in G^r(U)$ is defined similarly. If f is homogeneous of grade α, we have

$$(\partial x^i)^{-1}f\partial=\sigma(\alpha-g(i),g(i))\partial f(\partial x^i)^{-1}.\tag{6.4}$$

The following are elaborations of Propositions 3.9 and 3.10.

Proposition 6.7: We have

$$\partial^2f(\partial x^i)^{-1}(\partial x^j)^{-1}=\sigma(g(i),g(j))\partial^2f(\partial x^j)^{-1}(\partial x^i)^{-1}.$$

Proof: Straightforward.

Proposition 6.8: Let f, h be proper G^r-functions on U and suppose h is homogeneous of grade α. Assume that $\Pi_{j\in I_1}A_{g(j)}\neq 0$. Then we have

$$\partial(fh)(\partial x^i)^{-1} = \sigma(g(i),\alpha)\partial f(\partial x^i)^{-1}\cdot h + f\cdot \partial h(\partial x^i)^{-1}. \tag{6.5}$$

Proof: When i is even, (6.5) is true by Proposition 3.10. Let i be odd. Since $\partial(\partial x^i)$ is additive, we may suppose $f(x,\xi) = a(x)\xi^M$ and $h(x,\xi) = b(x)\xi^K$ for some $M, K\subset I_1$. If $i\notin M\cup K$ or some $j(\neq i)\in M\cap K$, then the both sides of (6.5) are zero. If $i\in M\cap K$, then $fh = 0$ and the left-hand side is zero. The right-hand side of (6.5) is equal to

$$\sigma(g(i),\alpha)\epsilon_1 a(x)\xi^{M-\{i\}}b(x)\xi^K + \epsilon_2 a(x)\xi^M b(x)\xi^{K-\{i\}}, \tag{6.6}$$

where

$$\xi^M = \epsilon_1\xi^{M-\{i\}}\xi^i,\ \xi^K = \epsilon_2\xi^{K-\{i\}}\xi^i.$$

Here we have

$$\begin{aligned}\epsilon_2 a(x)\xi^M b(x)\xi^{K-\{i\}} &= \epsilon_2\epsilon_1\sigma(g(i),\alpha - g(i))a(x)\xi^{M-\{i\}}b(x)\xi^{K-\{i\}}\xi^i\\ &= \epsilon_1\sigma(g(i),-g(i))\sigma(g(i),\alpha)a(x)\xi^{M-\{i\}}b(x)\xi^K.\end{aligned}$$

Since i is odd, $\sigma(g(i),-g(i)) = -1$ and (6.6) turns out to be zero.

Finally assume $M\cap K = \varnothing$ and $i\in M\cup K$. Since f and h are proper and $\xi^M\xi^K\neq 0$ by assumption, $fh = \sigma(g(M),\alpha - g(K))a(x)b(x)\xi^M\xi^K$ is the standard expansion of fh. Therefore, if $i\in K$ (the case when $i\in M$ is similar), then

$$\partial(fh)(\partial x^i)^{-1} = \epsilon\sigma(g(M),\alpha - g(K))a(x)b(x)\xi^M\xi^{K-\{i\}}, \tag{6.7}$$

where $\xi^K = \epsilon\xi^{K-\{i\}}\xi^i$. Since $\partial f(\partial x^i)^{-1} = 0$ and $\partial h(\partial x^i)^{-1} = \epsilon b(x)\xi^{K-\{i\}}$, (6.7) is equal to the right-hand side of (6.5).

Proper G^∞-functions on U form a σ-commutative G-graded algebra and Proposition 6.8 asserts that $\partial(\partial x^i)^{-1}$ is a G-graded superderivation of this algebra.

Remark 6.9: The operation $\partial(\partial x^i)^{-1}$ is not a derivation on the algebra $G^\infty(U)$ of all G^∞-functions on U, even if A is a Grassmann algebra. In fact, let A be the Grassmann algebra generated by a single element v. Let $f(x,\xi) = v$ and $h(x,\xi) = \xi$. Then $fh = 0$, but $-\partial f(\partial\xi)^{-1}h + f\partial h(\partial\xi)^{-1} = v\neq 0$.

VII. BEREZIN INTEGRALS

The integer r is such that the soul of every point of X has nilpotency r. Let $U\subset X = A_I$ be a domain and let $f\in G^r(U)$. We suppose U is equal to its saturation $\overline{U}$, in particular U contains its body $b(U)$.

We say that $f(x,\xi)\in G^r(U)$ has a *compact support*, if for any ξ_0 the restriction of $g(x) = f(x,\xi_0)$ to the body $b(U)$ has a compact support. Here $G^r_c(U)$ denotes the set of G^r-functions on U with compact support. Let $f(x,\xi) = \Sigma_M f_M(x)\xi^M$ be the standard expansion of f. Then f has a compact support if and only if every $f_M(x)$, $x\in b(U)$ has a compact support. We call f *singular* if $f_M(x)$ is in $\mathrm{Ann}(\Pi_{j\in I_1}A_{g(j)})$, for any $x\in b(U)$ and $M\subset I_1$.

Definition 7.1: For $f\in G^r_c(U)$, the *(Berezin) integral* of f on U is defined as

$$\int_U f(x,\xi)dx\,d\xi = \int_{b(U)} f_{I_1}(x)dx, \tag{7.1}$$

where $f_{I_1}(x)$ is the top of f. We should note that the integration in the right-hand side of (7.1) is defined in Definition 4.5.

Lemma 7.2: Let U and V be domains in A_I. Let $z = (z^i)$ be a G^{r+1}-mapping from V to A_I and $y = (y^i)$ be a G^{r+1}-mapping from U to V. Then $\Delta(z/x) - \Delta(z/y)\Delta(y/x)$ is a singular function in x, where the Jacobian is understood to be made from the canonical derivatives.

The proof of this lemma which is an elaboration of Proposition 5.2 is given in Ref. 14 and is omitted here. It is easily seen that the product of a singular function and any function is singular. Moreover, the integral of a singular function vanishes because its top is zero. Therefore we have the following lemma by Lemma 7.2.

Lemma 7.3: In the same situation as in Lemma 7.2 suppose $y(U) = V$ and $z(V) = W$ a domain in A_I. Let f be a G^r- function on W with a compact support. Then we have

$$\int_U f(z(y(x)))\Delta\left(\frac{z}{x}\right)dx = \int_U f(z(y(x)))\Delta\left(\frac{z}{y}\right)\Delta\left(\frac{y}{x}\right)dx.$$

Theorem 7.4: Let I be a G-set and U be a saturated domain in $X = A_I$. Let $y = (y^k,\eta^l)$ be a G^{r+1}-mapping of U to X and suppose that y is one to one and $\Delta(y/x)\neq 0$ on U. Let $V = y(U)$ and $f\in G^r_c(V)$. Then we have

$$\int_U f(y(x,\xi),\eta(x,\xi))\Delta\left(\frac{y}{x}\right)dx\,d\xi = \epsilon\int_V f(y,\eta)dy\,d\eta, \tag{7.2}$$

where $\epsilon = 1$ or $\epsilon = -1$ according to whether the Jacobian of the C^{r+1}-mapping $b(y)$ from $b(U)$ to $b(X)$ is positive or negative.

Proof: From (1) of Proposition 5.5, V is also saturated and the right-hand side of (7.2) makes sense. Since the given change of variables is decomposed into the following two types, Lemma 7.3 assures that it suffices to prove the assertion in each case separately:

(1) $y^k = y^k(x,\xi)$ and $\eta^l = \xi^l$,

(2) $y^k = x^k$ and $\eta^l = \eta^l(x,\xi)$.

The case (1) can be still broken up into the following subcases:

(1.1) $y^k = y^k(x)$ and $\eta^l = \xi^l$,

(1.2) $y^1 = x^1 + a(x)\xi^K$, $y^k = x^k$, for $k\neq 1$ and $\eta^l = \xi^l$, where $k(\neq\varnothing)\subset I_1$.

The case (1.1) can be reduced to the following two cases:

(1.1.1) $y^k = y^k(x)\in b(A)$, for $x\in b(U)$ and $\eta^l = \xi^l$,

(1.1.2) $y^1 = x^1 + a(x)$ with $a(x)^2 = 0$, $y^k = x^k$, for $k\neq 1$ and $\eta^l = \xi^l$.

The case (2) can be reduced to the following two cases:

(2.1) $y^k = x^k$ and $\eta^l = \sum_j a^l_j(x)\xi^j$, (2.2) $y^k = x^k,\ \eta^1 = \xi^1 + a(x)\xi^K,\quad \eta^l = \xi^l,$

for $l \neq 1$, where $K \subset I_1$ and $K \neq \{1\}$.

Let $f(y,\eta) = \Sigma_M f_M(y)\eta^M$ be the standard expansion of f. In the case (1.1.1), $y = y(x)$ is essentially the change of variables only on the body (Proposition 4.6), and so we have only to prove in other cases, where $b(U) = b(V)$ and $\epsilon = 1$ hold.

Case (1.1.2): We have $\Delta(y/x) = 1 + \partial a(\partial x^1)^{-1}$. Since the top of $f(y(x,\xi),\ \eta(x,\xi))\Delta(y/x)$ is equal to $P_{I_1}(f_{I_1}(y(x,\xi))\Delta(y/x))$ on the body $b(U)$, the following equalities hold:

$$\int_U f(y(x,\xi),\eta(x,\xi))\Delta\left(\frac{y}{x}\right) dx\, d\xi$$
$$= \int_{b(U)} P_{I_1}[f_{I_1}(x^1 + a(x),x^2,\ldots,x^p)$$
$$\times(1 + \partial a(\partial x^1)^{-1}(x))]dx$$
$$= P_{I_1}\int_{b(U)}\{f_{I_1}(x) + \partial f_{I_1}(\partial x^1)^{-1}(x)\cdot a(x)$$
$$+ f_{I_1}(x)\cdot\partial a(\partial x^1)^{-1}(x)$$
$$+ \partial f_{I_1}(\partial x^1)^{-1}(x)\cdot a(x)\partial a(\partial x^1)^{-1}(x)\}dx. \quad (7.3)$$

Since $a(x)^2 = 0$, we have $a(x)\cdot\partial a(\partial x^1)^{-1}(x) = 0$ by Proposition 3.10. Moreover by Proposition 4.7 we have

$$\int_{b(U)}(\partial f_{I_1}(\partial x^1)^{-1}(x)\cdot a(x)$$
$$+ f_{I_1}(x)\cdot\partial a(\partial x^1)^{-1}(x))dx$$
$$= \int_{b(U)}\partial(f_{I_1}(x)a(x))(\partial x^1)^{-1}\,dx = 0.$$

Thus (7.3) is equal to

$$P_{I_1}\int_{b(U)} f_{I_1}(x)dx$$
$$= \int_{b(U)} f_{I_1}(x)dx = \int_{b(V)} f_{I_1}(y)dy.$$

Case (1.2): We have

$$\Delta(y/x) = 1 + \sigma(g(1),g(K))\partial a(\partial x^1)^{-1}(x)\xi^K.$$

Noting the grade of $\partial a(\partial x^1)^{-1}(x)\xi^K$ is zero, we can calculate as follows:

$$\int_U f(y(x,\xi),\eta(x,\xi))\Delta\left(\frac{y}{x}\right) dx\, d\xi$$
$$= \int_U \sum_M \{f_M(x) + \partial f_M(\partial x^1)^{-1}(x)\cdot a(x)\xi^K\}\xi^M$$
$$\times(1 + \sigma(g(1),g(K))\partial a(\partial x^1)^{-1}(x)\xi^K)dx\, d\xi$$
$$= \int_U \{f_{I_1}(x)\xi^{I_1} + \partial f_{\bar K}(\partial x^1)^{-1}(x)\cdot a(x)\xi^K\xi^{\bar K}$$
$$+ \sigma(g(1),g(K))f_{\bar K}(x)\xi^{\bar K}\partial a(\partial x^1)^{-1}(x)\xi^K\}dx\, d\xi$$
$$= P_{I_1}\int_{b(U)}\{f_{I_1}(x) + \epsilon_1\cdot\partial f_{\bar K}(\partial x^1)^{-1}(x)\cdot a(x)$$
$$+ \epsilon_1\cdot\sigma(g(1),g(K))f_{\bar K}(x)\partial a(\partial x^1)^{-1}(x)\}dx, \quad (7.4)$$

where $\bar K = I_1 - K$ and $\xi^K\xi^{\bar K} = \epsilon_1\xi^{I_1}$. By Proposition 3.10 we have

$$\partial(f_{\bar K}\cdot a)(\partial x^1)^{-1}(x)$$
$$= \sigma(g(1),g(1) - g(K))\partial f_{\bar K}(\partial x^1)^{-1}(x)\cdot a(x)$$
$$+ f_{\bar K}(x)\cdot\partial a(\partial x^1)^{-1}(x),$$
$$= \sigma(g(K),g(1))[\partial f_{\bar K}(\partial x^1)^{-1}(x)\cdot a(x)$$
$$+ \sigma(g(1),g(K))f_{\bar K}(x)\partial a(\partial x^1)^{-1}(x)].$$

Hence by Proposition 4.7 we see that (7.4) is equal to

$$\int_{b(U)} f_{I_1}(x)dx = \int_{b(V)} f_{I_1}(y)dy.$$

Case (2.1): Let $M = (a^l_j(x))$, then $\Delta(y/x) = (\det M)^{-1}$ and $\eta^{I_1} = (\det M)\xi^{I_1}$ from Definitions 3.3 and 3.9 of Ref. 4. Hence

$$\int_U f(y(x,\xi),\eta(x,\xi))\Delta\left(\frac{y}{x}\right) dx\, d\xi$$
$$= \int_U f_{I_1}(x)\eta^{I_1}(\det M)^{-1}dx\, d\xi$$
$$= \int_U f_{I_1}(x)\xi^{I_1}dx\, d\xi = \int_V f(y,\eta)dy\, d\eta.$$

Case (2.2): First suppose $1\in K$ and let $K' = K - \{1\}$. Then we have

$$\Delta(y/x) = [1 + \sigma(g(1),g(K'))a(x)\xi^{K'}]^{-1}$$
$$= 1 - \sigma(g(1),g(K'))a(x)\xi^{K'}$$

and the following equalities:

$$\int_U f(y(x,\xi),\eta(x,\xi))\Delta\left(\frac{y}{x}\right) dx\, d\xi$$
$$= \int_U \left\{\sum_{1\notin M} f_M(x)\xi^M + \sum_{1\in M}(f_M(x)\xi^M\right.$$
$$\left. + f_M(x)a(x)\xi^K\xi^{M'})\right\}$$
$$\times[1 - \sigma(g(1),g(K'))a(x)\xi^{K'}]dx\, d\xi$$
$$= \int_U \{f_{I_1}(x)\xi^{I_1} + f_{\bar K'}(x)a(x)\xi^K\xi^{\bar K}$$
$$- \sigma(g(1),g(K'))f_{\bar K'}(x)\xi^{\bar K'}a(x)\xi^{K'}\}dx\, d\xi, \quad (7.5)$$

where $M' = M - \{1\}$, $\bar K = I_1 - K$, and $\bar K' = I_1 - K'$. Since the grade of $a(x)\xi^{K'}$ is zero we have

$$\xi^{\bar K'}a(x)\xi^{K'} = a(x)\xi^{K'}\xi^{\bar K'} = \sigma(g(K'),g(1))a(x)\xi^K\xi^{\bar K},$$

and (7.5) becomes

$$\int_{b(U)} f_{I_1}(x)dx = \int_{b(V)} f_{I_1}(y)dy.$$

Next suppose $1\notin K$. Then $\Delta(y/x) = 1$ and

$$\int_U f(y(x,\xi),\eta(x,\xi))\Delta\left(\frac{y}{x}\right) dx\, d\xi$$
$$= \int_U \left\{\sum_{1\notin M} f_M(x)\xi^M + \sum_{1\in M}(f_M(x)\xi^M\right.$$
$$\left. + f_M(x)a(x)\xi^K\xi^{M'})\right\}dx\, d\xi$$
$$= \int_{b(U)} f_{I_1}(x)dx = \int_{b(V)} f_{I_1}(y)dy.$$

The proof is complete.

Proposition 7.5: Let f and h be G^{r+1}-functions and assume that h has a compact support and a grade α. Then we have

$$\int_U f(x)\cdot\partial h(\partial x^i)^{-1}(x)dx\,d\xi$$
$$= -\sigma(g(i),\alpha)\int_U \partial f(\partial x^i)^{-1}\cdot h(x)dx\,d\xi. \qquad (7.6)$$

Proof: When i is even, the equality

$$\partial(fh)(\partial x^i)^{-1} = f\cdot\partial h(\partial x^i)^{-1} + \sigma(g(i),\alpha)\partial f(\partial x^i)^{-1}\cdot h \qquad (7.7)$$

holds. Since the top of fh has a compact support, the integral of the left-hand side of (7.7) is zero and (7.6) follows.

Next let i be odd. By the additivity of integration we may suppose that $f(x) = a(x)\xi^M$ and $h(x) = b(x)\xi^K$. If $i\notin M\cap K$ then the tops of the both integrands in (7.6) are zero. If $i\in M\cap K$ then the right-hand side of (7.7) is zero by the proof of Proposition 6.8, and (7.6) follows.

VIII. CONCLUDING REMARKS FOR FURTHER STUDIES

In the present paper we have developed differential and integral calculus on generalized superspaces, which will be a basis for deeper analysis of superfields. Except for Sec. II, we restrict the base field $\mathbf{F}$ to be the real field. An important and pressing problem is to extend our theory to the complex case.

Let A be the algebra of supernumbers over $\mathbf{R}$ and let $A_{\mathbf{C}} = A\otimes_{\mathbf{R}}\mathbf{C}$. No difficulty arises when we only extend the range of a function $f\in A^U$ to $A_{\mathbf{C}}$, where U is a domain in the *real* superspace X over A. A function $g\in A_{\mathbf{C}}^U$ is right differentiable at $x_0\in X$ with respect to x^i if there is a constant $a\in A_{\mathbf{C}}$ such that

$$g(x_0^1,\ldots,x_0^{i-1},x_0^i+y,x_0^{i+1},\ldots,x_0^{p+q})$$
$$= g(x_0) + ay + o(\|y\|).$$

The integral of g is also defined in the same manner as we did in Secs. IV and VII.

To extend the domain U to a domain in complex superspace is not quite trivial. We need to introduce a suitable involution in the algebra of complex supernumbers. This will be discussed as one of the main themes of our next paper.

In this paper integration is only defined for functions with compact support. For further studies of superfields it is important to consider integration for rapidly decreasing functions. It will be also treated in another paper.[14]

The consistency of integration proved in Sec. VII makes it possible to define integration for differential forms on (generalized) supermanifolds. We have a plan to write a paper about supermanifolds and differential forms on them.

[1]P. Fayet and S. Ferrara, Phys. Rep. C **32**, 249 (1977); B. S. DeWitt, *Supermanifolds* (Cambridge U.P., Cambridge, 1984).

[2]D. A. Leites, Usp. Mat. Nauk. **35**, 3 (1980); A. Rogers, J. Math. Phys. **21**, 1352 (1980); A. Jadczyk and K. Pilch, Commun. Math. Phys. **78**, 373 (1981); J. Hoyos, M. Quiros, J. Ramíres Mittelbrunn, and F. J. de Urries, J. Math. Phys. **25**, 847 (1984); C. P. Boyer and S. Gitler, Trans. Am. Math. Soc. **285**, 241 (1984).

[3]M. Scheunert, J. Math. Phys. **20**, 712 (1979).

[4]Y. Kobayashi and S. Nagamachi, J. Math. Phys. **25**, 3367 (1984).

[5]V. K. Agrawala, Hadronic J. **4**, 444 (1981); H. S. Green and P. D. Jarvis, J. Math. Phys. **24**, 1731 (1983); M. Scheunert, *ibid.* **24**, 2658, 2671, 2681 (1983).

[6]B. Mitra and K. C. Tripathy, J. Math. Phys. **25**, 2550 (1984).

[7]H. S. Green, Phys. Rev. **90**, 270 (1953); Y. Ohnuki and S. Kamefuchi, *Quantum Field Theory and Parastatistics* (Univ. Tokyo P., Tokyo, 1982).

[8]A. Rogers, J. Math. Phys. **21**, 1352 (1980).

[9]A. Rogers, J. Math. Phys. **26**, 385 (1985).

[10]F. A. Berezin, *Introduction to Algebra and Analysis with Anticommuting Parameters* (Moscow U. P., Moscow, 1983).

[11]J. Hoyos, M. Quiros, J. Ramíres Mittelbrunn, and F. J. de Urries, J. Math. Phys. **25**, 841 (1984).

[12]A. Jadczyk and K. Pilch, Commun. Math. Phys. **78**, 373 (1981).

[13]The same formulation was given by Rogers in Ref. 8 as z expansions for G^∞-functions in the supercommutative case.

[14]Y. Kobayashi and S. Nagamachi, Lett. Math. Phys. **11**, 293 (1986).

Linearization and Painlevé property of Liouville and Cheng equations

K. M. Tamizhmani and M. Lakshmanan
Department of Physics, Bharathidasan University, Tiruchirapalli-620 023, India

(Received 25 October 1985; accepted for publication 30 April 1986)

It is demonstrated that the Liouville equation and the Cheng equation (describing a chemical reaction) are free from movable critical manifolds and possess the Painlevé property. The associated linearizing transformations and general solutions follow naturally from the Painlevé analysis.

I. INTRODUCTION

In recent times, the Painlevé property for partial differential equations (PDE's) has drawn much attention[1–7] either in the original sense of Ablowitz, Ramani, and Segur[1] or in the generalized form of Weiss, Tabor, and Carnevale.[4] In the latter approach, a PDE is said to possess the Painlevé property if its solution can be expressed as a single-valued expansion about a noncharacteristic singular manifold[4] $\varphi(x,t) = 0$. In the present note, we discuss the Painlevé property and Bäcklund transformations for the Liouville[8] and Cheng[9] equations. Consequently, the linearizing transformations and the general solutions are shown to follow automatically.

II. THE LIOUVILLE EQUATION

First, we consider the Liouville equation[8]

$$u_{xt} - e^u = 0. \tag{1}$$

Under the transformation

$$u = \log V, \tag{2}$$

(1) becomes

$$VV_{xt} - V_x V_t - V^3 = 0. \tag{3}$$

We now look for solutions of (3) in the form

$$V = \varphi^\alpha \sum_{j=0}^{\infty} V_j \varphi^j, \tag{4}$$

where V_j and φ are analytic functions of (x,t) in a neighborhood of the singularity manifold $\varphi(x,t) = 0$, and α is a negative integer, to be determined. Inserting $V \approx V_0 \varphi^\alpha$ in (3), by leading-order analysis, we find that $\alpha = -2$ and

$$V_0 = 2\varphi_x \varphi_t. \tag{5}$$

Substituting (4) in (3) and equating the coefficient of φ^{j-6} to zero, we get

$$(j+1)(j-2) V_0 V_j \varphi_x \varphi_t = 0, \tag{6}$$

and so the resonance values are $j = -1, 2$. The resonance $j = -1$ corresponds to the arbitrariness of the manifold $\varphi = 0$. At $j = 1$, we find that

$$V_1 = -2\varphi_{xt}. \tag{7}$$

Also, we observe that at $j = 2$ the resulting equation is satisfied identically and so V_2 is arbitrary. Thus (1) possesses the Painlevé property.

By cutting off the series (4) at the constant level term ($V_j = 0, j \geqslant 3$), we find the Bäcklund transformation in the form

$$V = (2\varphi_x \varphi_t/\varphi^2) - (2\varphi_{xt}/\varphi) + V_2, \tag{8}$$

where both V and V_2 satisfy (3) and

$$\begin{aligned}&\varphi_x \varphi_t^2 V_{2x} + \varphi_x^2 \varphi_t V_{2t} \\ &\quad + (2\varphi_x \varphi_t \varphi_{xt} - \varphi_x^2 \varphi_{tt} - \varphi_{tt}^2 \varphi_{xx}) V_2 \\ &\quad - \varphi_t \varphi_{xx} \varphi_{xtt} - \varphi_{xt} \varphi_{xx} \varphi_{tt} - \varphi_x \varphi_t \varphi_{xxtt} = 0,\end{aligned} \tag{9}$$

$$\begin{aligned}&\varphi_x \varphi_t V_{2xt} - 2\varphi_{xt} \varphi_t V_{2x} - \varphi_x \varphi_{tt} V_{2x} \\ &\quad - 2\varphi_x \varphi_{xt} V_{2t} - \varphi_t \varphi_{xx} V_{2t} \\ &\quad + (\varphi_{xx} \varphi_{tt} + 2\varphi_{xxt} \varphi_t - 4\varphi_{xt}^2 + 2\varphi_x \varphi_{xtt}) V_2 \\ &\quad - 3\varphi_x \varphi_t V_2^2 + 2\varphi_{xt} \varphi_{xxtt} - 2\varphi_{xxt} \varphi_{xtt} = 0,\end{aligned} \tag{10}$$

and

$$\varphi_{xtt} V_{2x} + \varphi_{xxt} V_{2t} + 3\varphi_{xt} V_2^2 - \varphi_{xt} V_{2xt} - \varphi_{xxtt} V_2 = 0 \tag{11}$$

hold.

When we consider the vacuum solution[4] $V_2 = 0$, it easily can be shown that the admissible solution to (9)–(11) is given by

$$\varphi_{xt} = 0. \tag{12}$$

Consequently by (2), (8) becomes

$$u = \log(2\varphi_x \varphi_t/\varphi^2), \tag{13}$$

where φ satisfies the linearized wave equation (12). From (12) and (13), we infer that the transformation (13) maps the solution of the linearized wave equation to a solution of the Liouville equation. Moreover, the general solution to (12) is $\varphi(x,t) = g(x) + h(t)$, where g and h are arbitrary functions of x and t, respectively, and so (13) becomes

$$u = \log(2g_x h_t/(g+h)^2), \tag{14}$$

which is the known general solution of the Liouville equation[10] (1).

A similar analysis can be performed for the Dodd–Bullough (DB) equation $u_{xt} = e^u - e^{-2u}$, which can be rewritten as $VV_{xt} - V_x V_t - V^3 + 1 = 0$, using (2). The dominant behavior is given by $\alpha = -2$ and $V_0 = 2\varphi_x \varphi_t$. The resonances are again $j = -1, 2$ and we find that the DB equation possesses the Painlevé property. However, we notice that it does not admit the linearizing transformation in the sense discussed earlier, due to the constant term on the left-hand side.

 0022-2488/86/092257-02$02.50

III. A CHEMICAL REACTION EQUATION

Next, we consider the Cheng equation[9]

$$u_x = -auv, \quad v_t = bu_x, \tag{15}$$

where a and b are constants, corresponding to the dynamics of the photosensitive molecules when the light beam passes through them. We expand

$$u = \sum_{j=0}^{\infty} u_j \varphi^{j-1}, \quad v = \sum_{j=0}^{\infty} v_j \varphi^{j-1} \tag{16}$$

and find from the leading-order analysis that

$$u_0 = (1/ab)\varphi_t, \quad v_0 = (1/a)\varphi_x. \tag{17}$$

We further find that resonances occur in (16) at $j = -1, 1$. Substituting (16) in (15), and equating the coefficients of $(\varphi^{-1}, \varphi^{-1})$ to zero, we obtain two equations, the first one being

$$(\varphi_{xt}/ab) + \varphi_x u_1 + (\varphi_t/b)v_1 = 0, \tag{18}$$

while the second equation is identically zero. This implies that either the function u_1 or v_1 is arbitrary. Thus the system (15) possesses the Painlevé property.

As in the previous example, we find the Bäcklund transformation in the form

$$u = (1/ab)(\varphi_t/\varphi) + u_1, \quad v = (1/a)(\varphi_x/\varphi) + v_1, \tag{19}$$

where (u,v) and $(u_1 v_1)$ satisfy (15) and

$$\frac{1}{\varphi}\left(\frac{\varphi_{xt}}{ab} + \varphi_x u_1 + \frac{\varphi_t}{b} v_1\right) = 0, \tag{20}$$

which is identically satisfied because of (18). Now, considering the vacuum solutions $u_1 = 0$ and $v_1 = 0$, from (20), we arrive at exactly the same linearized wave equation (12). This allows us to write the general traveling wave solutions of (15) from (19) in the form

$$u = \frac{1}{ab}\left(\frac{h_t}{g(x) + h(t)}\right), \quad v = \frac{1}{a}\left(\frac{g_x}{g(x) + h(t)}\right), \tag{21}$$

where $g(x)$ and $h(t)$ are arbitrary functions discussed in Sec. II. Solution (21) is indeed the general solution derived by Cheng[9] from his analysis for (15).

Here, we have constructed the linearizing transformations and solutions of (1) and (15) in a rather simple and straightforward manner, from the Painlevé analysis.

ACKNOWLEDGMENTS

KMT wishes to thank the Department of Atomic Energy, Government of India, for providing a Senior Research Fellowship and the work reported here of ML forms part of a research project sponsored by the Indian National Science Academy.

[1]M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715, 1006 (1980).
[2]M. Lakshmanan and P. Kaliappan, J. Math. Phys. **24**, 795 (1983).
[3]J. B. McLeod and P. J. Olver, SIAM J. Math. Anal. **14**, 488 (1983).
[4]J. Weiss, M. Tabor, and G. Carnevale, J. Math. Phys. **24**, 522 (1983).
[5]J. Weiss, J. Math. Phys. **24**, 1405 (1983); **25**, 2226 (1984).
[6]K. M. Tamizhmani and R. Sahadevan, J. Phys. A **18**, L1067 (1985).
[7]R. Sahadevan, K. M. Tamizhmani, and M. Lakshmanan, "Painlevé analysis and integrability of coupled nonlinear Schrödinger equations," J. Phys. A **19**, (1986).
[8]A. K. Pogrebkov and M. K. Polivanov, Sov. J. Part. Nucl. **14**, 450 (1983).
[9]H. Cheng, Stud. Appl. Math. **70**, 183 (1984).
[10]L. V. Ovsiannikov, *Group Analysis of Differential Equations*, translation edited by W. F. Ames (Academic, New York, 1982), p. 112.

The anomaly structure of theories with external gravity

L. Bonora, P. Pasti, and M. Tonin
Dipartimento di Fisica "G. Galilei," Via Marzolo, 8, 35100 Padova, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Padova, Padova, Italy

(Received 11 November 1985; accepted for publication 30 April 1986)

The cohomology problem of the overall local symmetry group of theories with external gravity, including diffeomorphisms, local Lorentz, and gauge transformations, is studied, in order to determine all possible anomalies. To this end the nontrivial cohomology classes of the coupled system of two coboundary operators are classified in the abstract. Using this result and a technical assumption the nontrivial cohomology classes of the coboundary operator associated with diffeomorphisms are determined. These possible anomalies split in any dimension into two distinct families. Both are calculated (the second only in four dimensions). Using known results about gauge and local Lorentz anomalies, the possible anomalies of the overall local symmetry group are determined.

I. INTRODUCTION

The recent rise of interest in anomalies has produced a better knowledge of the geometrical and algebraic origins of chiral anomalies[1] and has permitted us to assimilate the known gravitational anomalies[2–4] partly to Lorentz gauge anomalies[5] and partly to Weyl anomalies.[6]

Unfortunately we do not have yet a general argument that suggests that these are the only possible anomalies.

In this paper we tackle the problem of finding all possible anomalies of the group of diffeomorphisms in a theory including (external) gravity. We use cohomological (or consistency) methods,[7–10] so that possible anomalies are represented by nontrivial cohomology classes of the coboundary operator corresponding to general coordinate transformations. This problem has been recently investigated also by Bandelloni.[11]

Before we turn to the results of this paper we must clearly specify our program. For the sake of manageability the differential space the coboundary operator acts upon has been restricted in this paper to be the space F of local functionals, which are integrated polynomials of the connections, gauge fields, matter fields, vielbeins, and inverse vielbeins (see the exact definition at the beginning of Sec. IV). This space includes all known actions and anomalies, while it excludes Bardeen–Zumino-type actions.[5]

In this way we are able to solve completely the cohomology problem, i.e., to find all the nontrivial cohomology classes in the space F. Of course we cannot exclude the existence of other nontrivial cohomology classes not contained in F.

With the above limitation in mind, our results are specified by Theorem 4.1, which holds in any space-time dimension and gives the first family of nontrivial cohomology classes, and by (5.40), i.e., the second family. The latter is calculated in four dimensions.

These results are quite general since we include in our analysis all fields appearing in a theory with gravity, i.e., the vielbeins, inverse vielbeins, connections, gauge fields, and matter fields; moreover, they are valid also in the presence of torsion and for nonmetric connections. The only limitation concerns the fact that we have assumed the conventional point of view of a field theory defined on a chart rather than globally defined in some manifold, therefore we have not worried about the objects we have used being globally defined.

The previous results, however, must be considered as an intermediate step in our program. Indeed, (1) we must verify whether we can eliminate the nontrivial cocycles we have found by means of Bardeen–Zumino-type counterterms, as is the case for Eq. (5.40) and for the first family above; and (2) we must compare diffeomorphisms with other symmetries of the theory, since, by subtracting counterterms from the quantum action, we may violate other symmetries.

In fact, in the present papers we do not limit ourselves to studying the cohomology of diffeomorphisms. We study the cohomology of the most general (local) symmetry group of a given theory with (external) gauge and gravitational fields. This is motivated by the trivial fact that the (one-loop) Ward identities corresponding to symmetry transformations of the classical theory can be written as a unique Ward identity

$$\Sigma\hat{\Gamma} = O(\hbar),$$

where $\hat{\Gamma}$ is the vertex generating functional and Σ is the sum of all functional operators generating the symmetry transformations. Once the group parameters become FP ghosts (see Sec. II), endowed with the specific transformation laws that express the associativity of the overall symmetry group transformations, Σ becomes nilpotent, and we can study the relevant cohomology. This cohomology accounts for the relations among cocycles generated by distinct Ward identities. For the purpose of studying this coupled cohomology we have proved Theorem 3.1, which classifies the nontrivial cohomology classes of the sum of two coboundary operators in terms of the cocycles of each.

It turns out that our results on diffeomorphisms together with well-known results about Lorentz and gauge anomalies are sufficient to determine the nontrivial cohomology classes of Σ. From the latter we can extract all possible anomalies of a gauge theory coupled to gravity in four di-

 0022-2488/86/092259-12$02.50

mensions. They are simply the usual gauge anomalies and a mixed U(1)-gravitational anomaly, which can take several different forms (see the end of Sec. VII). In particular, the cocycle given in Eq. (5.40), which is nontrivial in the space F, is not an anomaly in four dimensions since it can be canceled by a Bardeen–Zumino counterterm.

The article is arranged as follows. Section II is devoted to definitions, notations, and conventions. In Sec. III we prove the above-mentioned classification theorem and analyze a few general consequences. In Sec. IV we determine the first family of anomalies of the diffeomorphisms; Sec. V is devoted to the second family. In Sec. VI we comment on the results found in Secs. IV and V. Finally in Sec. VII we determine the cohomology of the overall symmetry group, including diffeomorphisms and local Lorentz and gauge transformations.

II. NOTATIONS AND CONVENTIONS

For any local symmetry group S of a classical theory with infinitesimal parameters $\epsilon\lambda^\alpha(x)$ $(\alpha = 1,...,N)$ we shall introduce a coboundary operator Σ_S in the following way. Let us denote by $\varphi_r \to \varphi_r + \delta_S\varphi_r(\varphi,\lambda)$ the local infinitesimal transformation on the generic field φ_r of the theory. This transformation is operated by the functional operator

$$\overline{\Sigma}_S = \int_x \delta_S\varphi_r \frac{\delta}{\delta\varphi_r},$$

where the summation over r is understood. Let us consider now the λ^α's as anticommuting fields (FP ghosts) and endow them with a transformation property: $\lambda^\alpha(x) \to \lambda^\alpha(x) + \delta_S\lambda^\alpha(x)$. Then we introduce the operator

$$\Sigma_S = \int_x \delta_S\lambda^\alpha(x)\frac{\delta}{\delta\lambda^\alpha(x)} + \overline{\Sigma}_S. \tag{2.1}$$

There is a choice of $\delta_S\lambda^\alpha(x)$ such that

$$\Sigma_S^2 = 0. \tag{2.2}$$

This is what we refer to as the coboundary operator corresponding to the symmetry S. The choice of $\delta_S\lambda^\alpha$, which renders Σ_S nilpotent, is dictated by the geometry of the group S (see Ref. 9): the λ^α are to be assimilated to the Maurer–Cartan form on S and the $\delta_S\lambda^\alpha$'s express the Maurer–Cartan equation. When we want to indicate explicitly the dependence of Σ_S on the ghost λ^α, we shall write Σ_S^λ instead of Σ_S.

The coboundary operator Σ_S can be defined also when the symmetry S is global. In that case the FP ghosts are constant anticommuting parameters.

The invariance of the classical action I under the symmetry S is

$$\Sigma_S I = 0. \tag{2.3}$$

For the quantized theory it implies the Ward identity (WI)

$$\Sigma_S\widehat{\Gamma} = \hbar\Delta_S + 0(\hbar^2), \tag{2.4}$$

where $\widehat{\Gamma}$ is the vertex generating functional. For the purposes of this paper it is enough to limit ourselves to one-loop order and to the case of external ghosts. Here Δ_S is a local functional of the fields and their derivatives; it is linear in λ^α and of mass dimension equal to the space-time dimension n, if $\dim\lambda^\alpha$ is determined in such a way that $\dim\Sigma_S = 0$. Moreover, as a consequence of Eqs. (2.2) and (2.4), Δ_S satisfies the consistency condition

$$\Sigma_S\Delta_S = 0, \tag{2.5}$$

i.e., Δ_S is a cocycle of Σ_S. In general, Δ_S is a sum of independent cocycles $\Delta_S^{(i)}$. If there exists a local functional $C^{(i)}$ independent of λ^α such that

$$\Delta_S^{(i)} = \Sigma_S C^{(i)}, \tag{2.6}$$

$\Delta_S^{(i)}$ can be adsorbed through a redefinition of $\widehat{\Gamma}$. If

$$\Delta_S^{(i)} \neq \Sigma_S C, \tag{2.7}$$

for any local functional C, the symmetry is broken at the quantum level.

In quantum field theories where all fields have canonical dimensions >0, Δ_S and $\Delta_S^{(i)}$, $C^{(i)}$, and C in Eqs. (2.5), (2.6), and (2.7) are integrals of local polynomials of the fields and their derivatives. In theories containing fields with vanishing canonical dimensions this fact is not as obvious. Anyway, in this paper, we shall investigate only the cocycles $\Delta_S^{(i)}$ of the relevant coboundary operator Σ_S in the space of P-functionals. For theories involving gravity, P-functionals will be defined at the beginning of Sec. IV. Here $\Delta_S^{(i)}$ is a coboundary if it satisfies Eq. (2.6), where $C^{(i)}$ is a local P-functional independent of λ^α (local action). If Eq. (2.7) holds for any local action C, $\Delta_S^{(i)}$ is a nontrivial cocycle (which we call an a-cocycle).

It is among these a-cocycles that we must look for anomalies according to the program explained in the Introduction.

As anticipated by the terms we have used, Eqs. (2.5)–(2.7) set up a cohomological problem. Indeed a differential space[12] is defined by the couple formed by the vector space of P-functionals and by the nilpotent operator Σ_S acting on it.[11]

The problem consists of determining the cohomology space, that is, the set of nontrivial cohomology classes, each being identified by a-cocycles that differ from one another by coboundaries. In this paper we shall be concerned with this problem for the coboundary operators listed below, leaving aside the question of whether there exists a renormalized $\widehat{\Gamma}$ that actually generates the a-cocycles. We shall comment on this question in Sec. VI.

We are interested in the cohomology problem for the Lie algebras of the group of diffeomorphisms D with parameters $\xi^l(x)$, the group of local Lorentz transformations L on the tangent space with parameters $u_a^b(x)$ and a generic gauge group G with parameters $\lambda^\alpha(x)$. According to the above recipe the parameters become anticommuting ghost fields with transformation laws:

$$\delta_D\xi^l = \xi^p\partial_p\xi^l, \tag{2.8}$$

$$\delta_L u_b^a = -u_b^c u_c^a, \tag{2.9}$$

$$\delta_G\lambda^\alpha = -\tfrac{1}{2}f^{\alpha\beta\gamma}\lambda^\beta\lambda^\gamma, \tag{2.10}$$

where the $f^{\alpha\beta\gamma}$ are the structure constants of the gauge group. The field transformation laws are the usual ones. We only write down the transformation law for D relative to the affine connection Γ_{mn}^l for the sake of stating the sign conventions

$$\delta_D \Gamma^l_{mn} = \bar{\delta}_D \Gamma^l_{mn} + \partial_m \partial_n \xi^l, \qquad (2.11)$$

where $\bar{\delta}_D$ denotes the transformation of Γ considered as a covariant tensor. Then, according to Eq. (2.1), we define the coboundary operators Σ_D, Σ_L, and Σ_G and denote by Δ_D, Δ_L, and Δ_G the relative cocycles.

In theories that, classically, are simultaneously D-, L-, and G-invariant, the relevant coboundary operator is not each operator Σ_D, Σ_L, or Σ_G separately, but the sum

$$\Sigma = \Sigma_D + \Sigma_L + \Sigma_G. \qquad (2.12)$$

Thus Σ is nilpotent,

$$\Sigma^2 = 0, \qquad (2.13)$$

provided that the ghosts u^b_a and λ^α are considered as scalar fields (with weight 0) under D. The appropriate modification of Σ_D is understood.

The cohomology of Σ is determined on the basis of the cocycles of the various operators with the addition of important restrictive conditions (see the next section).

For later use we introduce also the subgroups D_A and $\bar{D}_A$ of D. Here D_A is the group GL(4R) whose infinitesimal parameters are obtained by specializing $\xi^m(x)$ to $\xi^m_A(x) = x^l \alpha^m_l$, where α^m_l are generic constants and $\bar{D}_A$ is the subgroup SL(4R) of D_A, whose infinitesimal parameters satisfy the traceless condition $\alpha^l_l = 0$. We define correspondingly two coboundary operators Σ_A and $\bar{\Sigma}_A$ by promoting α^m_l to anticommuting (constant) ghosts with the transformation law

$$\Sigma_A \alpha^m_l = \alpha^n_l \alpha^m_n \qquad (2.14)$$

and the same law for $\bar{\Sigma}_A$. Since a theory that is classically D-invariant is also classically D_A-invariant ($\bar{D}_A$-invariant), it makes sense and proves useful to define a coupled coboundary operator $\Sigma_D + \Sigma_A$ ($\Sigma_D + \bar{\Sigma}_A$), which is indeed nilpotent provided that we postulate the following obvious cross-transformation laws:

$$\Sigma_A \xi^m = x^l \alpha^n_l \partial_n \xi^m + \xi^n \alpha^m_n, \quad \Sigma_D \alpha^m_l = 0, \qquad (2.15)$$

and the same for $\bar{\Sigma}_A$.

III. THE COUPLED COHOMOLOGY PROBLEM

While studying the cohomology of Σ_D and that of Σ [Eq. (2.12)], we are faced with the problem of finding the a-cocycles of a coboundary operator that is the sum of two coboundary operators in terms of the cocycles of the latter. Therefore, in this section, we solve this problem in general.

Let S and R be two symmetries and Σ_S and Σ_R be the relative coboundary operators. Let us define the mixed coboundary operator $\Sigma_S + \Sigma_R$, $(\Sigma_S + \Sigma_R)^2 = 0$. The cocycles of $\Sigma_S + \Sigma_R$ have the form $\Delta_S + \Delta_R$, where Δ_S (Δ_R) is a cocycle of Σ_S (Σ_R). Indeed the consistency condition

$$(\Sigma_S + \Sigma_R)(\Delta_S + \Delta_R) = 0 \qquad (3.1)$$

implies, in particular, $\Sigma_S \Delta_S = 0$ and $\Sigma_R \Delta_R = 0$. Given Δ_S, a Δ_R satisfying Eq. (3.1), if it exists, is defined up to cocycles $\bar{\Delta}_R$ of Σ_R satisfying the condition $\Sigma_S \bar{\Delta}_R = 0$. We remark that such $\bar{\Delta}_R$'s are cocycles of $\Sigma_S + \Sigma_R$. The only (up to coboundaries) Δ_R such that $\Delta_S + \Delta_R$ belongs to a definite cohomology class of $\Sigma_S + \Sigma_R$ is called the *R-partner* of Δ_S.

We shall say that an a cocycle of Σ_S is *admissible* if it has an R-partner.

For the sake of conciseness, let us introduce the concept of an S-symmetry-preserving a-cocycle of Σ_R, briefly, an S-a-cocycle of Σ_R. By this we mean a cocycle Δ_R of Σ_R that satisfies the following condition:

$$\Sigma_S \Delta_R = 0 \quad \text{and} \quad \Delta_R \neq \Sigma_R C, \quad \text{for any } C \text{ s.t. } \Sigma_S C = 0, \qquad (3.2)$$

S-a-cocycles and a-cocycles do not, in general, coincide.

Now we classify all the a-cocycles of $\Sigma_S + \Sigma_R$ according to the characteristics of the cocycles Δ_S.

Case (1). Δ_S is an a-cocycle of Σ_S: $\Sigma_S \Delta_S = 0$, $\Delta_S \neq \Sigma_S C$, $\forall C$.

(1a) $\Sigma_R \Delta_S = 0$. In this case Δ_S is an a-cocycle of $\Sigma_S + \Sigma_R$.

(1b) $\Sigma_R \Delta_S \neq 0$. Among all these Δ_S we look for linear combinations $\tilde{\Delta}_S$ for which an R-partner $\tilde{\Delta}_R$ exists. If such a $\tilde{\Delta}_S$ exists then $\tilde{\Delta}_S + \tilde{\Delta}_R$ is an a-cocycle of $\Sigma_S + \Sigma_R$.

Case (2). Δ_S is a coboundary of Σ_S: $\Delta_S = \Sigma_S C$. In this case an R-partner Δ_R certainly exists. However it may occur that $\Sigma_R C = \Delta_R + \bar{\Delta}_R$. It follows that $(\Sigma_S + \Sigma_R)\bar{\Delta}_R = 0$. Therefore $\bar{\Delta}_R$ is a cocycle of $(\Sigma_S + \Sigma_R)$ and in particular $\Sigma_S \bar{\Delta}_R = 0$.

(2a) $\bar{\Delta}_R = 0$. Then $\Delta_S + \Delta_R$ is a coboundary of $\Sigma_S + \Sigma_R$.

(2b) $\bar{\Delta}_R \neq 0$ and $\bar{\Delta}_R \neq \Sigma_R C$, $\forall C$. Then both $\Delta_R + \Delta_S$ and $\bar{\Delta}_R$ are a-cocycles of $\Sigma_S + \Sigma_R$. They belong to the same cohomology class, for $\Delta_S + \Delta_R + \bar{\Delta}_R = (\Sigma_S + \Sigma_R)C$.

(2c) $\bar{\Delta}_R \neq 0$ and $\bar{\Delta}_R = \Sigma_R \bar{C}$ for some $\bar{C}$.

(2c1) $\Sigma_S \bar{C} = 0$. Then both $\Delta_S + \Delta_R$ and $\bar{\Delta}_R$ are coboundaries of $\Sigma_S + \Sigma_R$.

(2c2) $\Sigma_S \bar{C} \neq 0$. $\Delta_S + \Delta_R$ and $\bar{\Delta}_R$ are a-cocycles of $\Sigma_S + \Sigma_R$ belonging to the same cohomology class.

We summarize the results obtained as follows.

Theorem 3.1: The nontrivial cohomology classes of $\Sigma_S + \Sigma_R$ are uniquely determined (1) by the linear combinations of a-cocycles of Σ_S that admit an R-partner and by the relative R-partners, and (2) by the S-a-cocycles of Σ_R (with vanishing S-partners).

The next corollary follows immediately.

Corollary 3.2: If Σ_S does not have nontrivial cohomology classes, the only admissible a-cocycles of Σ_R are the S-a-cocycles.

This corollary expresses, in general, the relation between absence of anomalies in a given WI and exactness of the corresponding symmetry. Otherwise stated it says that disregarding S-symmetry violating P-functionals and local actions implies only the loss of coboundaries of $\Sigma_S + \Sigma_R$.

Remark 3.3: In Theorem 3.1, we can reverse the role of R and S. Then it is easy to realize that the nontrivial cohomology classes of $\Sigma_S + \Sigma_R$ fall into three different groups: the first is determined by the S-a-cocycles of Σ_R (with vanishing S-partners), the second by the R-a-cocycles of Σ_S (with vanishing R-partners) and the third by a-cocycles of Σ_S whose R-partners are a-cocycles of Σ_R.

Remark 3.4: Let us particularize the above results to the case when R is a subgroup of S. We denote by $\lambda^a_R(x)$ the restriction of the infinitesimal parameters $\lambda^a(x)$ of S to R. It is useful to consider R and S as independent groups. Thus when promoting λ and λ_R to ghost fields we shall consider them as independent. In the few cases we are interested in, in the next sections, it turns out that if we set

$$\Sigma_S^\lambda \lambda_R = 0, \quad \Sigma_R^{\lambda_R}\lambda = \Sigma_S^{\lambda_R}\lambda, \tag{3.3}$$

we can define the coupled coboundary operator $\Sigma_S + \Sigma_R$. Now, we can apply Theorem 3.1, but, in this case, we can get more independent information. Indeed, if $\Delta_S = \int \lambda \mathfrak{a}$ is a cocycle of Σ_S, then $\Delta_R = \int \lambda_R \mathfrak{a}$ is a cocycle of Σ_R and

$$(\Sigma_S + \Sigma_R)(\Delta_S + \Delta_R)$$
$$= (\Sigma_S + \Sigma_R)\int (\lambda + \lambda_R)\mathfrak{a} = \Sigma_S^{\lambda+\lambda_R}\int(\lambda + \lambda_R)\mathfrak{a} = 0.$$

Therefore both Δ_S and Δ_R are admissible. This remark will be used in the next section.

IV. THE COHOMOLOGY OF DIFFEOMORPHISMS: THE FIRST FAMILY OF a-COCYCLES

First of all we specify the vector space where the coboundary operator Σ_D operates. It is the vector space F of P-functionals, that is, of integrated local polynomials of the vielbeins, inverse vielbeins, connection, and all the other fields involved and their derivatives, with canonical dimensions equal to the space-time dimension n ($\dim \xi^m = -1$).

Since $\Sigma_D F \subset F$, the couple (Σ_D, F) is a differential space in which the cohomology problem can be consistently defined. We remark that F includes all the known actions used in field theories and all known anomalies. However F does not include Bardeen–Zumino-like actions [see Eqs. (6.3) and (7.3)]. This feature implies that nontrivial cohomology classes of Σ_D in F cannot be identified immediately with anomalies, as explained in the Introduction. However, it has the advantage that we can solve the cohomology problem completely. This is what we want to show in this and the next section.

A word of caution is in order (even though it is rather obvious): with our procedure we miss possible nontrivial cohomology classes not belonging to F.

Before proceeding we need another specification about locality. On a general ground, we should start from the cohomology of Σ_D in the space of P-functionals containing also powers of the coordinates x^m. However, as is shown in Appendix A, rigid translations do not have nontrivial cohomology classes. Due to Corollary 3.2, this allows us to study the cohomology in the space of local P-functional F. Therefore, from now on we shall refer to the local cohomology.

We shall proceed by analyzing first the cohomology of Σ_A and $\overline{\Sigma}_A$. This allows us, in general, to delimit the possible form of the cocycles of Σ_D and, in particular, to find a set of a-cocycles of Σ_D that is present in any space-time dimension.

This is the content of the present section and the relevant results are summarized in Theorem 4.1. The remaining part of the analysis is restricted to four dimensions and is carried out in Sec. V.

The cohomology of Σ_A is analyzed in Appendix B. The most general admissible form of a cocycle is

$$\Delta_A = \int \alpha^n_m \left(\sum_{r,s} \sum_{\substack{k_1\cdots k_r \\ l_1\cdots l_s}} C^{l_1\cdots l_s}_{k_1\cdots k_r}\; {}^w\mathfrak{a}^{mk_1\cdots k_r}_{nl_1\cdots l_s} \right), \tag{4.1}$$

where the ${}^w\mathfrak{a}^{mk_1\cdots k_r}_{nl_1\cdots l_s}$ are local polynomials of the fields and their derivatives of $D_{\overline{A}}$ weight w where the world indices not appearing explicitly are understood to be saturated, and $\int \equiv \int d^D x$. The $C^{l_1\cdots l_s}_{k_1\cdots k_r}$ are numerical coefficients not containing constant tensors. Moreover, the summations over r and s are finite. There exist many solutions of the consistency conditions $\Sigma_A \Delta_A = 0$. They correspond to coboundaries of Σ_A except for the case $r = s = 0$.

In this case

$$\Sigma_A \mathfrak{a}_n{}^m = x^l \alpha_l{}^p \partial_p \mathfrak{a}_n{}^m + \alpha_n{}^l \mathfrak{a}_l{}^m - \alpha_l{}^m \mathfrak{a}_n{}^l + w\alpha_l{}^l \mathfrak{a}_n{}^m,$$

where w is the D_A-weight of $\mathfrak{a}^m_n$ (not to be confused with the D-weight).

The consistency condition is

$$\Sigma_A \Delta_A = \Sigma_A \int \alpha_m{}^n \mathfrak{a}_n{}^m$$
$$= \int (-\alpha_m{}^p \alpha_p{}^n \mathfrak{a}_n{}^m + (w-1)\alpha_l{}^l \alpha_m{}^n \mathfrak{a}_n{}^m) = 0,$$

which can only be satisfied if $\mathfrak{a}^m_n = \delta^m_n \mathfrak{a}$. It is easy to see that if $w \neq 1$, Δ_A is a coboundary, while for $w = 1$, it is an a-cocycle. Therefore the a-cocycles of Σ_A have the form

$$\Delta_A = \alpha_l{}^l \int \mathfrak{a}, \tag{4.2}$$

where $\mathfrak{a}$ is a scalar with $w = 1$ under D_A.

The first consequence is that $\overline{\Sigma}_A$ does not have a-cocycles. Therefore, from Corollary 3.2, it follows that the a-cocycles of Σ_D must be $\overline{D}_A$-preserving. Of course, the same holds also for the a-cocycles of Σ_L and Σ_G. Moreover, since the operator analogous to $\overline{\Sigma}_A$ in the tangent space does not have a-cocycles either, we are entitled from now on to restrict our study to P-functionals in which both world and tangent space indices are completely saturated. However the cohomology of Σ_A tells us much more, provided that we remember that D_A is to be considered as a subgroup of D. Therefore the relevant a-cocycles are those admissible with respect to the cohomology $\Sigma_A + \Sigma_D$.

Now we want to relate these cocycles to the a-cocycles of Σ_D in order to extract information about the latter. To this end let us write the generic cocycle of Σ_D as $\Delta_D = \int \xi^m(x)\mathfrak{b}_m(x)$. Then the condition $\overline{\Sigma}_A \Delta_D = 0$ implies

$$\overline{\Sigma}_A \mathfrak{b}_m = x^l \alpha_l{}^n \partial_n \mathfrak{b}_m + \alpha_m{}^n \mathfrak{b}_n.$$

As a consequence

$$\Sigma_A \mathfrak{b}_m = x^l \alpha_l{}^n \partial_n \mathfrak{b}_m + \alpha_m{}^n \mathfrak{b}_n + w\alpha_l{}^l \mathfrak{b}_m, \tag{4.3}$$

where w is the D_A-weight of $\mathfrak{b}_m$. Now let us consider the coupled cohomology $\Sigma_A + \Sigma_D$. We apply Theorem (3.1) with D_A in the place of S and D in the place of R. Then we have two sets of a-cocycles of $\Sigma_A + \Sigma_D$.

(1) The first set is determined by the a-cocycles Δ_A of Σ_A given by Eq. (3.2) that admit a D-partner. Let us look for a D-partner $\Delta_D = \int \xi^m \mathfrak{b}_m$ for each of these Δ_A's:

$$(\Sigma_A+\Sigma_D)\Big(\alpha_l^l\int\mathfrak{a}+\int\xi^m\mathfrak{b}_m\Big)$$
$$=-\alpha_l^l\Sigma_D\int\mathfrak{a}+(w-1)\alpha_l^l\int\xi^m\mathfrak{b}_m=0, \tag{4.4}$$

where Eqs. (2.15) and (4.3) have been used. We remark that Eq. (4.4) can be satisfied only when $w=1$. For, if $w\neq 1$, it implies that acting with Σ_D on $\int\mathfrak{a}$ ($\mathfrak{a}$ has D_A-weight 1), one gets a P-functional with weight $w\neq 1$, which is impossible (see the beginning of the next section). Therefore $w=1$ and

$$\Sigma_D\int\mathfrak{a}=0. \tag{4.5}$$

(2) The second set is determined by the D_A-a-cocycles of Σ_D,

$$0=(\Sigma_A+\Sigma_D)\int\xi^m\mathfrak{b}_m=(w-1)\alpha_l^l\int\xi^m\mathfrak{b}_m, \tag{4.6}$$

which can only be satisfied if $w=1$.

Conclusion: The admissible a-cocycles of Σ_D have D_A-weight 1 and the admissible a-cocycles of Σ_A must satisfy Eq. (4.5).

Now, if $\int\xi^m\mathfrak{b}_m$ is an a-cocycle of Σ_D, it is an admissible cocycle w.r.t. $\Sigma_A+\Sigma_D$, due to Remark 3.4. This implies that its D_A-weight w is 1. On the other hand, $\int x^l\alpha_l^m\mathfrak{b}_m$ is certainly a cocycle of Σ_A (see Remark 3.4), i.e.,

$$\Sigma_A\int x^l\alpha_l{}^m\mathfrak{b}_m$$
$$=-\int x^l\alpha_l{}^m\alpha_m{}^p\mathfrak{b}_p+(1-w)\int x^l\alpha_l{}^m\alpha_p{}^p\mathfrak{b}_m=0.$$

Since $w=1$,

$$\int x^l\alpha_l{}^m\alpha_m{}^p\mathfrak{b}_p=0. \tag{4.7}$$

This equation can be satisfied only if

$$\mathfrak{b}_m=-\partial_m\mathfrak{b}+\partial_{p_1}\partial_{p_2}\mathfrak{b}_m^{p_1p_2}, \tag{4.8}$$

where $\mathfrak{b}$ is not itself a derivative. Again, since $\alpha_l^l\int\mathfrak{b}$ must be an admissible cocycle it must satisfy Eq. (4.5). This equation implies, in particular, that $\mathfrak{b}$ has D-weight 1. Thus we have proven the following theorem.

Theorem 4.1: The most general a-cocycle of Σ_D has the form

$$\Delta_D=\int(\partial_m\xi^m\mathfrak{b}+\partial_{p_1}\partial_{p_2}\xi^m\mathfrak{b}_m^{p_1p_2}),$$

where $\mathfrak{b}$ is a D-scalar density with D-weight 1 (and is not a derivative), and $\mathfrak{b}_m^{p_1p_2}$ is a D_A-tensor with D_A-weight 1 with explicit form to be further determined. The $\mathfrak{b}$ and $\mathfrak{b}_m^{p_1p_2}$ define the *first* and *second family* of a-cocycles of Σ_D.

V. THE COHOMOLOGY OF DIFFEOMORPHISMS: THE SECOND FAMILY OF a-COCYCLES

So far the analysis has been carried out without any dimensional restriction. However, in order to derive the explicit form of the a-cocycles of the second family, we must find all the solutions of the consistency equation of the form $\int\partial_{p_1}\partial_{p_2}\xi^m\mathfrak{b}_m^{p_1p_2}$. Although the method we are going to use can certainly be generalized to n dimensions, in this paper we shall limit ourselves to four dimensions.

It is well known that a general connection splits into a metric connection plus a nonmetricity tensor. However, in order to find a very general solution (applicable also to gauge theories) we forget in this section the relation between metric and connection and treat them as uncorrelated fields. We shall study the implications of this relation in the next section. With this proviso the only possible a-cocycles we find are specified by Eqs. (5.10), (5.26), and (5.40). In the next section we shall show that the cocycles of Eqs. (5.10) and (5.26) are coboundaries (in F), while Eq. (5.40) is not. For the sake of simplicity we drop the cautionary adjective possible throughout this section.

First let us write $\mathfrak{b}_m^{p_1p_2}$ in such a form as to exhibit an unambiguous separation between the covariant and the noncovariant parts. To this end let us remark that any D_A-tensor with D_A-weight w can be written as a polynomial of Γ's (Γ denotes Γ^l_{mn}) with D-covariant coefficient of D-weight w. For example, whenever we come across a covariant derivative D_m applied to a D-tensor T with D-weight w, we introduce the derivative $\overline{D}_m=D_m-w\Gamma^l_{ml}$; $\overline{D}_mT$ is D-covariant with D-weight w. In conclusion, in four dimensions, we can write $\mathfrak{b}_m^{p_1p_2}$ as follows:

$$\mathfrak{b}=B_1+\Gamma B_2+\Gamma\Gamma B_3+\Gamma\Gamma\Gamma B_4$$
$$+\partial\Gamma B_5+\Gamma\,\partial\Gamma B_6+\partial\partial\Gamma B_7, \tag{5.1}$$

where all the indices are understood (for instance, $\partial\Gamma B_5$ means $\partial_{s_3}\Gamma^r_{s_1s_2}B_{5m}^{p_1p_2s_1s_2s_3}{}_r$) and the B_i ($i=1,\ldots,7$) are covariant D-tensors of D-weight 1. Since we wish to discriminate between covariant and noncovariant parts, we split Γ as follows: $\Gamma=\widetilde{\Gamma}+T$, where $\widetilde{\Gamma}$ is the symmetric part of Γ and T is the torsion tensor. Since T is covariant we can absorb it in a redefinition of the coefficients B_i. We suppose that this has already been done and that Γ appearing in Eq. (5.1) and in the remaining part of this section is actually $\widetilde{\Gamma}$.

Moreover, consider, for example, the term B_5. If it is antisymmetric in, say, s_1 and s_3, then using the definition of the curvature, we can absorb this term into B_1 and B_3. Therefore, in order to avoid ambiguities (and without loss of generality), we assume B_5 to be completely symmetric in s_1, s_2, and s_3. The same remark applies to B_6 and B_7.

As a second preliminary step, let us split Σ_D into two functional operators:

$$\Sigma_D=\Sigma_D^c+\widehat{\Sigma}_D. \tag{5.2}$$

Here Σ_D^c, when applied to a monomial of the fields and their derivatives with given weight (and with saturated or unsaturated world indices), transforms it as Σ_D would if the monomial were a covariant tensor with the same indices and the same weight. Then $\widehat{\Sigma}_D$ is defined by Eq. (5.2). In particular, we have

$$\widehat{\Sigma}_D\Gamma_m{}^l{}_n=\partial_m\partial_n\xi^l,$$
$$\widehat{\Sigma}_D\,\partial_m\partial_n\xi^l=0, \tag{5.3}$$
$$\widehat{\Sigma}_D\,\partial_l\xi^m=-\partial_l\xi^p\partial_p\xi^m,\quad\text{etc.}$$

One can prove that

$$\widehat{\Sigma}_D^2=0. \tag{5.4}$$

In general, unlike Σ_D, $\hat{\Sigma}_D$ does not commute with the operation of differentiation, except in special cases. For instance, it does commute with the exterior derivative d applied to forms without external indices (see below).

It is also convenient to use the language of differential forms. Let us denote by $Q^i_j, P^i_j, R^i_j, \ldots$ j-forms whose components are polynomials of the fields and their derivatives with ghost number i. Then

$$\Sigma_D \int Q^1_4 = \hat{\Sigma}_D \int Q^1_4, \quad \Sigma_D \int Q^0_4 = \hat{\Sigma}_D \int Q^0_4. \tag{5.5}$$

This shows that cocycles and a-cocycles of Σ_D are cocycles and a-cocycles, respectively, of $\hat{\Sigma}_D$ and vice versa. Therefore from now on we shall use only $\hat{\Sigma}_D$.

Let us return to the cocycles of the second family:

$$\Delta_D = \int \partial_{p_1} \partial_{p_2} \xi^m \mathfrak{b}^{p_1 p_2}_m \equiv \int Q^1_4, \quad \hat{\Sigma}_D \Delta_D = 0. \tag{5.6}$$

This implies that there exists a three-form Q^2_3 such that

$$\hat{\Sigma}_D Q^1_4 = dQ^2_3, \tag{5.7a}$$

where d represents the exterior derivative. Applying $\hat{\Sigma}_D$ to this equation and using Eq. (5.4), we get $\hat{\Sigma}_D dQ^2_3 = 0$. As previously stated in this and the following cases, $\hat{\Sigma}_D d = d\hat{\Sigma}_D$. Therefore using the local Poincaré lemma[9] we can conclude that

$$\hat{\Sigma}_D Q^2_3 = dQ^3_2, \tag{5.7b}$$

for some Q^3_2. Similarly

$$\hat{\Sigma}_D Q^3_2 = dQ^4_1, \tag{5.7c}$$

$$\hat{\Sigma}_D Q^4_1 = dQ^5_0, \tag{5.7d}$$

$$\hat{\Sigma}_D Q^5_0 = 0, \tag{5.7e}$$

for suitable Q^i_j.

It is easier to solve Eqs. (5.7) with high ghost number than those with low ghost number. Therefore whenever possible, we try to reduce our problem of finding the a-cocycles of $\hat{\Sigma}_D$ of the type Q^1_4 to the problem of finding the a-cocycles of $\hat{\Sigma}_D$ of the type Q^2_3, Q^3_2, etc. The method essentially consists in looking for a complete and reasonably simple classification of the solutions of Eq. (5.7a). As we shall see, a simple classification is provided by the solutions of $\hat{\Sigma}_D(Q^1_4 - dP^1_3) = 0$ and $\hat{\Sigma}_D(Q^2_3 - dP^2_2) = 0$, which are specified by Theorems 5.1 and 5.4 below. What is left out from this classification can be determined easily through a direct calculation. Both theorems are divided into a part I and a part II. Although only the first parts are essential for our final results, we prove part II for reasons that will be clear shortly.

First selection: Let us consider the cocycles Δ_D of Eq. (5.6) that satisfy the equation

$$\hat{\Sigma}_D(Q^1_4 - dP^1_3) = 0, \tag{5.8}$$

for some three-form P^1_3. Let us separate the possible P^1_3's into two classes A and B, according to the following distinction: a polynomial or form is class A if it contains only $\partial\partial\xi$ or higher derivatives of ξ, while it is class B if it contains at least one factor ξ or $\partial\xi$.

Theorem 5.1:

Part I: The cocycle Δ_D of Eq. (5.6) is a coboundary if and only if Q^1_4 satisfies Eq. (5.8) for some class A P^1_3.

Part II: If Q^1_4 satisfies Eq. (5.8) for a class B P^1_3, then either it is one of the coboundaries of part I or it is an a-cocycle having the form

$$\partial_m \xi^m \partial_l K^l, \tag{5.9}$$

where K^l is a noncovariant polynomial tensor. The only example, up to coboundaries, is the following:

$$\Delta^{(1)}_D = \int \mathrm{Tr}(\Lambda)\mathrm{Tr}(d\Gamma)\mathrm{Tr}(d\Gamma). \tag{5.10}$$

Here we have introduced the matrix notation[13]

Λ for the 4×4 matrix $\Lambda^n_m = \partial_m \xi^n$,

Γ for the 4×4 matrix 1-form $\Gamma^n_m = \Gamma^n_{lm}\, dx^l$.

Proof of part I: If P^1_3 is class A, then the necessary condition is obvious since Δ_D being a coboundary means

$$Q^1_4 = \hat{\Sigma}_D P^0_4 + dP^1_3,$$

for suitable P^0_4 and P^1_3. Then Eq. (5.8) is a consequence of Eq. (5.4). That the condition is sufficient is proven in Appendix C.

Proof of part II: In order to prove the second part of the theorem (and for later use) it is convenient to introduce the following notation: for any p-form ω in an n-dimensional space, whose components $\omega_{i_1\cdots i_p}$ are polynomials of the field and their derivatives, let us call dual polynominal tensors the quantities

$$\tilde{\omega}^{i_1\cdots i_{n-p}} = \epsilon^{i_1\cdots i_n}\omega_{i_{n-p+1}\cdots i_n}, \tag{5.11}$$

where $\epsilon^{i_1\cdots i_n}$ is the constant completely antisymmetric tensor (with weight 1). When a metric is defined, the $\tilde{\omega}^{i_{p+1}\cdots i_n}$ are the components of the form dual to ω,[14] with all the indices raised. To $d\omega$ there corresponds a dual polynomial tensor

$$\tilde{\alpha}^{i_1\cdots i_{n-p-1}} = \epsilon^{i_1\cdots i_n}\partial_{i_{n-p}}\omega_{i_{n-p+1}\cdots i_n} = \partial_{i_{n-p}}\tilde{\omega}^{i_1\cdots i_{n-p}}. \tag{5.12}$$

This correspondence is obviously one to one.

Using the dual tensors and remembering the form of Q^1_4 [Eq. (5.6)], we can rewrite Eq. (5.8) for a class B P^1_3 as

$$\hat{\Sigma}_D(\partial_{p_1}\partial_{p_2}\xi^m \mathfrak{b}^{p_1 p_2}_m - \partial_l(\xi^i C^l_i + \partial_m \xi^i D^{ml}_i)) = 0, \tag{5.13}$$

where $\mathfrak{b}^{p_1p_2}_m$ is given by Eq. (5.1) and C and D have analogous expressions. Due to the transformations (5.3), Eq. (5.13) implies in particular

$$\partial_l C^l_i = 0. \tag{5.14}$$

It follows that

$$C^l_i = \partial_p C^{lp}_i, \tag{5.15}$$

with C^{lp}_i antisymmetric in l and p (this is nothing but the Poincaré lemma applied to the dual tensors). Then

$$\partial_l(\xi^i C^l_i) = \partial_p(\partial_l \xi^i C^{lp}_i). \tag{5.16}$$

Therefore C^l_i can be absorbed into D^{ml}_i and we can drop it in Eq. (5.13). The latter implies, now, that either

$$\partial_l D^{ml}_i = 0 \tag{5.17}$$

or

$$D^{ml}_i = \delta^m_i K^l. \tag{5.18}$$

Arguing as above, we can absorb D_l^{ml} into $b_m^{p_1p_2}$ if Eq. (5.17) is satisfied. If Eq. (5.18) holds, we have

$$\partial_{p_1}\partial_{p_2}\xi^m\hat{\Sigma}_D b_m^{p_1p_2} - \partial_l\,\partial_m\xi^m\hat{\Sigma}_D K^l - \partial_m\xi^m\hat{\Sigma}_D\,\partial_l K^l = 0. \tag{5.19}$$

That is,

$$\hat{\Sigma}_D\partial_l K^l = 0, \quad b_m^{pl} = \delta_m^p B^l, \quad \hat{\Sigma}_D(B^l - K^l) = 0, \tag{5.20}$$

which means that $\partial_l K^l$ and $B^l - K^l$ are invariant and covariant, respectively, with weight 1. Here $B^l - K^l$ identifies a coboundary (see part I), and $B^l = K^l$, with $\partial_l K^l$ invariant, represents a distinct solution,

$$\int \partial_l\,\partial_m\xi^m K^l = -\int \partial_m\xi^m\,\partial_l K^l, \tag{5.21}$$

which is a coboundary (already found in part I) if K^l is covariant, since

$$\int \partial_l\partial_m\xi^m K^l = \hat{\Sigma}_D\int \Gamma_{lm}^m K^l,$$

and is an *a*-cocycle if K^l is noncovariant.

Corollary 5.2: Q_4^1 is a coboundary if and only if the corresponding Q_3^2 defined by (5.7a) satisfies the condition

$$Q_3^2 = \hat{\Sigma}_D P_3^1 + dP_2^2, \tag{5.22}$$

for some class *A* P_3^1 and some P_2^2.

Equations (5.7a) and (5.8) imply $d(Q_3^2 - \hat{\Sigma}_D P_3^1) = 0$ for a class A P_3^1. The Poincaré lemma gives Eq. (5.22). Vice versa, by applying the exterior differential to Eq. (5.22), one gets $dQ_3^2 = d\hat{\Sigma}_D P_3^1 = \hat{\Sigma}_D Q_4^1$, due to Eq. (5.7a). Then from Theorem 5.1, Q_4^1 is a coboundary.

Another important limitation comes from the following lemma.

Lemma 5.3: Q_3^2 defined by Eq. (5.7a) can be written in a class A form, that is in a form bilinear either in $\partial\partial\xi$ or in $\partial\partial\xi$ and $\partial\partial\partial\xi$.

Let us consider the dual tensor $\tilde{Q}_3^2$ of Q_3^2. The general form of $\tilde{Q}_3^2$ is

$$\begin{aligned}\tilde{Q}_3^2 = {}& \xi\xi F_1 + \xi\partial\xi F_2 + \xi\partial\partial\xi F_3 + \xi\partial\partial\partial\xi F_4 + \xi\partial\partial\partial\partial\xi F_5 \\ & + \xi\partial\partial\partial\partial\partial\xi F_6 + \partial\xi\partial\xi F_7 + \partial\xi\partial\partial\xi F_8 + \partial\xi\partial\partial\partial\xi F_9 \\ & + \partial\xi\partial\partial\partial\partial\xi F_{10} + \partial\partial\xi\partial\partial\xi F_{11} + \partial\partial\xi\partial\partial\partial\xi F_{12}.\end{aligned} \tag{5.23}$$

Here the indices have been dropped. For example, $\xi\xi F_1$ stands for $\xi^i\xi^j F_{1_{ij}}^{\,l}$. Now, as a consequence of Eqs. (5.1), (5.6), and (5.4), dQ_3^2, is class A.

Then, in particular, $\partial_l F_{1_{ij}}^{\,l} = 0$, which implies, through the Poincaré lemma, that $F_{1_{ij}}^{\,l} = \partial_m F_{1_{ij}}^{\,lm}$, where $F_{1_{ij}}^{\,lm}$ is antisymmetric in *l,m*:

$$\partial_l(\xi^i\xi^j F_{1_{ij}}^{\,l}) = \partial_m((\partial_l\xi^i\xi^j + \xi^i\partial_l\xi^j)\,F_{1_{ij}}^{\,lm}). \tag{5.24}$$

Therefore F_1 can be absorbed into F_2. We can do the same for F_i, $i = 1,...,10$, and find that either they vanish (since $\partial F_1 = 0$, for $i = 6,10$, implies $F_i = 0$ because they have dimension 0), or they can be absorbed into F_{11} and F_{12}.

Second selection: We repeat almost step by step the above procedure for forms with a higher ghost number. Let us consider the solutions of the equation

$$\hat{\Sigma}_D(Q_3^2 - dP_2^2) = 0, \tag{5.25}$$

for some two-form P_2^2, where Q_3^2 is given by Eq. (5.7a) and specified by Lemma 5.3.

Theorem 5.4:

Part I: Q_3^2 identifies a coboundary if and only if it satisfies Eq. (5.25) for a class A P_2^2.

Part II: If Q_3^2 satisfies Eq. (5.25) for a class B P_2^2, then either it is one of the coboundaries of part I or the problem of identifying Q_3^2 can be formally reduced to the cohomology problem for diffeomorphisms in two dimensions. As a result of this analysis we identify another nontrivial cohomology class in four dimensions. As a representative we may choose

$$\Delta_D^{(2)} = \int \mathrm{Tr}(\Lambda\, d\Gamma)\mathrm{Tr}(d\Gamma). \tag{5.26}$$

Proof of part I: If Q_3^2 corresponds to a coboundary, then Eq. (5.22) holds. Since Q_3^2 is specified by Lemma 5.3 and P_3^1 is class A, then dP_2^2 is class A, too. Arguing as in Lemma 5.3 we can prove that P_2^2 is class A. Applying $\hat{\Sigma}_D$ to Eq. (5.22) we get Eq. (5.25) and the necessary condition is proven. The sufficiency proof is a simplified version of the analogous proof in part I of Theorem 5.1.

Proof of part II: Arguing as in part II of Theorem 5.1 we can write Eq. (5.25) in the form (a few details are given in Appendix D)

$$\hat{\Sigma}_D(\partial_{p_1}\partial_{p_2}\xi^m G_m^{p_1p_2r} - \partial_l(\partial_p\xi^m H_m^{plr})) = 0, \tag{5.27}$$

where *G* and *H* are class A polynomials. If $\partial_l H_m^{plr} = 0$, then *H* can be absorbed into *G*; therefore it can only correspond to a coboundary. If

$$H_m^{plr} = \delta_m^p H^{lr}, \tag{5.28}$$

then

$$\hat{\Sigma}_D(\partial_{p_1}\partial_{p_2}\xi^m G_m^{p_1p_2r} - \partial_l\,\partial_m\xi^m H^{lr}) = 0 \tag{5.29}$$

and

$$\hat{\Sigma}_D\,\partial_l H^{lr} = 0. \tag{5.30}$$

Equation (5.28) identifies only coboundaries (see part I) unless

$$G_m^{plr} = \delta_m^p G^{lr} \quad \text{and} \quad G^{lr} - H^{lr} = 0.$$

In this case the solutions are determined by Eq. (5.30), which, in terms of differential forms, is

$$\hat{\Sigma}_D dP_2^1 = d\hat{\Sigma}_D P_2^1 = 0, \tag{5.31}$$

provided that H^{lr} is the dual tensor of P_2^1. Equation (5.31) implies

$$\hat{\Sigma}_D P_2^1 = dP_1^2, \tag{5.32}$$

for some P_1^2. Equation (5.31) states a cohomology subproblem, which is formally equivalent to the problem studied in this section in two dimensions (formally, because we are actually dealing with objects defined in four dimensions). This subproblem can be easily solved (Appendix D). If P_2^1 is a coboundary of this subproblem, i.e., if

$$P_2^1 = \hat{\Sigma}_D R_2^0 + dR_1^1, \tag{5.33}$$

then $\hat{\Sigma}_D(d\partial_m\xi^m P_2^1) = 0$, since arguing as in Lemma 5.3 one can prove that R_1^1 equals $\partial\partial\xi$ times a covariant coefficient. Therefore, due to part I, P_2^1 defines a coboundary of the main problem. There are two distinct *a*-cocycles of the subproblem

$$(P_2^1)^{(1)} = \mathrm{Tr}(\Lambda)\mathrm{Tr}(d\Gamma),$$
$$(P_2^1)^{(2)} = \mathrm{Tr}(\Lambda\, d\Gamma). \tag{5.34}$$

The other a-cocycles differ from these by coboundaries. The first, when replaced into Eq. (5.27), is annihilated due to Eq. (5.28). The second corresponds to a $(P_2^2)^{(2)} = \mathrm{Tr}(\Lambda)\mathrm{Tr}(\Lambda\, d\Gamma)$, which cannot be set into such a form as to satisfy the hypotheses of part I of this theorem. Therefore it defines a nontrivial cohomology class of the main problem, a representative of which is the a-cocycle in Eq. (5.26).

What remains: What is left out by the previous analysis has to be looked for among the Q_4^1's such that the corresponding Q_2^3 defined by Eq. (5.7b) does not vanish.

Theorem 5.5: The tensor dual to Q_2^3 can be written in the form

$$\partial_{p_1}\Lambda^i_{p_2}\,\partial_{q_1}\Lambda^j_{q_2}\,\Lambda^k_r E^{p_1p_2q_1q_2r\ lm}_{i\ \ j\ \ k}, \tag{5.35}$$

where

$$E^{p_1p_2q_1q_2r\ lm}_{i\ \ j\ \ k} = \epsilon^{p_1q_1lm}\widetilde{E}^{p_2q_2r}_{i\ j\ k} \tag{5.36}$$

and $\widetilde{E}$ has weight 0 and is formed only with Kronecker δ's.

The proof is based on the fact that, since Q_3^2 is class A (Lemma 5.3), $\hat{\Sigma}_D Q_3^2$ also is. Using the same argument as in Lemma 5.3, we conclude that the dual tensor of Q_2^3 can be written in the form (5.35), where E is a covariant tensor with weight 1 and canonical dimension 0, antisymmetric in l and m and can be chosen antisymmetric under the exchange of the group of indices $\binom{p_1p_2}{i}$ and $\binom{q_1q_2}{j}$. Now, since the dual tensor associated with dQ_2^3 is

$$\partial_{p_1}\Lambda^i_{p_2}\,\partial_{q_1}\Lambda^j_{q_2}\,\partial_l\Lambda^k_r E^{p_1p_2q_1q_2r\ lm}_{i\ \ j\ \ k}$$
$$+ 2\partial_l\partial_{p_1}\Lambda^i_{p_2}\partial_{q_1}\Lambda^j_{q_2}\Lambda^k_r E^{p_1p_2q_1q_2r\ lm}_{i\ \ j\ \ k}$$
$$+ \partial_{p_1}\Lambda^i_{p_2}\,\partial_{q_1}\Lambda^j_{q_2}\,\Lambda^k_r\,\partial_l E^{p_1p_2q_1q_2r\ lm}_{i\ \ j\ \ k}, \tag{5.37}$$

and must be of class A, we have in particular $\partial_l E = 0$, which implies that E is a constant tensor, i.e., it is formed with Kronecker δ's and with the totally antisymmetric tensor ϵ^{mnlp}. Since also the second term on the right-hand side of Eq. (5.37) must vanish the only possible form of E is given by Eq. (5.36).

Theorem 5.5 brings our analysis to an end, since it is now very easy to classify all possible Q_2^3's. Up to total differentials (which correspond to coboundaries, due to Theorem 5.4), there are only three distinct possibilities:

$$Q_2^{3(1)} = \mathrm{Tr}(d\Lambda)\mathrm{Tr}(d\Lambda)\mathrm{Tr}(\Lambda),$$
$$Q_2^{3(2)} = \mathrm{Tr}(d\Lambda\, d\Lambda)\mathrm{Tr}(\Lambda), \tag{5.38}$$
$$Q_2^{3(3)} = \mathrm{Tr}(d\Lambda\, d\Lambda\, \Lambda).$$

We have

$$\hat{\Sigma}_D Q_2^{3(1)} = \hat{\Sigma}_D Q_2^{3(2)} = 0,$$
$$\hat{\Sigma}_D Q_2^{3(3)} = \tfrac{1}{3}\, d\,\mathrm{Tr}(d\Lambda\Lambda\Lambda\Lambda).$$

If $Q_2^{3(i)}$ $(i = 1,2,3)$ were to correspond to coboundaries, then from Eq. (5.7b) and Theorem 5.4 we would have

$$Q_2^{3(i)} = \hat{\Sigma}_D P_2^{2(i)} + dP_1^{3(i)}, \quad i = 1,2,3, \tag{5.39}$$

for suitable $P_1^{3(i)}$ and class A $P_2^{2(i)}$. However Eqs. (5.39) are not satisfied, in particular for $i = 1,2$ because of the clause that $P_2^{2(i)}$ must be class A. For $Q_2^{3(3)}$ we can easily determine $Q_1^{4(3)} = \frac{1}{3}\mathrm{Tr}(d\Lambda\Lambda\Lambda\Lambda)$ and $Q_0^{5(3)} = \frac{1}{15}\mathrm{Tr}(\Lambda\Lambda\Lambda\Lambda\Lambda)$ and see that were they to correspond to a coboundary, than there would exist a $P_0^{4(3)}$ such that $Q_0^{5(3)} = \hat{\Sigma}_D P_0^{4(3)}$. No such $P_0^{4(3)}$ exists.[15]

Due to Theorem 5.1 part I and Theorem 5.2 part I, $Q_2^{3(i)}$ $(i = 1,2,3)$ uniquely identify the only three distinct nontrivial cohomology classes belonging to the second family. One can easily see that $Q_2^{3(1)}$ and $Q_2^{3(2)}$ correspond to the cohomology class identified by $\Delta_D^{(1)}$ [Eq. (5.10)] and $\Delta_D^{(2)}$ [Eq. (5.26)], respectively, while a representative for the third cohomology class is

$$\Delta_D^{(3)} = \int \mathrm{Tr}\Big(d\Lambda\Big(d\Gamma\Gamma - \frac{1}{2}\Gamma\Gamma\Gamma\Big)\Big). \tag{5.40}$$

As one can see, only part I of Theorems (5.1) and (5.4) are strictly necessary to prove our result. However, part II of these theorems reveals the recursive character of the D-anomalies when dimensions increase. Beside the a-cocycle of the Adler–Bardeen type [Eq. (5.40)], we have other factorized a-cocycles that contain as factors lower-dimensional a-cocycles [we may consider $\mathrm{Tr}(\Lambda)$ a zero-dimensional a-cocycle]. The procedure we have presented suggests a clear pattern to generalize the results of this paper to n dimensions.

We recall that the construction of counterterms throughout this section was carried out by explicitly forgetting the relation between metric and connection. In this way we have found a result applicable also to gauge theories characterized by nonsemisimple Lie groups. Indeed if we replace Γ by the relevant gauge connection and Λ by the gauge ghost we can repeat almost verbatim the proof of this section. In this case the cocycles corresponding to $\Delta_D^{(1)}$, $\Delta_D^{(2)}$, and $\Delta_D^{(3)}$ are true anomalies, and this result together with a suitably adapted version of Sec. IV provides a uniqueness proof for anomalies in gauge theories.[16]

VI. DISCUSSION OF THE PREVIOUS RESULTS

Let us discuss about the candidates for a-cocycles $\Delta_D^{(i)}$ $(i = 1,2,3)$ of Eqs. (5.10), (5.26), and (5.40) in the light of the splitting

$$\Gamma^l_{mn} = \overline{\Gamma}^l_{mn} + N^l_{mn}, \tag{6.1}$$

where $\overline{\Gamma}^l_{mn}$ is a metric connection and N^l_{mn} is the nonmetricity tensor. When we insert the splitting (6.1) in $\Delta_D^{(1)}$ and $\Delta_D^{(2)}$ the pieces depending only on the symmetric part of $\overline{\Gamma}$ vanish and the remaining terms are coboundaries in F. For example, in $\Delta_D^{(2)}$ one of the surviving terms is $\int \mathrm{tr}(d\Lambda N)\mathrm{tr}(dN)$: it satisfies the consistency condition and, as is implicit from the theorems of the previous section, it is trivial: it is indeed generated by $\int \mathrm{tr}(\Gamma N)\mathrm{tr}(dN)$ [here N is the matrix one-form $(N)^l_n = N^l_{mn}\, dx^m$].

When we insert the splitting (6.1) in $\Delta_D^{(3)}$ we obtain many coboundaries depending on N. For example, the term linear in N is

$$\overline{\Delta}_D^{(3)} = \int_x \mathrm{Tr}\Big[d\Lambda\Big(d\overline{\Gamma}\,N + N\,d\overline{\Gamma} - \frac{1}{2}N\overline{\Gamma}\overline{\Gamma} - \frac{1}{2}\overline{\Gamma}N\overline{\Gamma} - \frac{1}{2}\overline{\Gamma}\overline{\Gamma}N\Big)\Big]$$

and

$$\hat{\Sigma}_D \Delta_D^{(3)} = 0, \quad \overline{\Delta}_D^{(3)} = \hat{\Sigma}_D \overline{C}^{(3)},$$

where

$$\overline{C}^{(3)} = \int_x \mathrm{Tr}\Big(\overline{\Gamma}\, d\overline{\Gamma}\, N + \overline{\Gamma} N\, d\overline{\Gamma} + \frac{3}{2} N\overline{\Gamma}\overline{\Gamma}\overline{\Gamma}\Big).$$

We are left with the term depending only on $\overline{\Gamma}$, which does not vanish in this case and is nontrivial in F.

In the previous section we have separated Γ into its symmetric part $\tilde{\Gamma}$ and torsion T and shown that only the symmetric part enters the cocycles $\Delta_D^{(i)}$, $i = 1,2,3$, implying that cocycles (of the second family) containing T must be coboundaries. This is more evidence of the statement that when we add to Γ^l_{mn} any covariant tensor Z^l_{mn} of weight zero we get new cocycles from the old ones: the differences between the old and new cocycles are coboundaries. We can use this fact to split Γ into the Christoffel symbols plus suitable tensors: up to coboundaries we are therefore left with $\Delta_D^{(3)}$ constructed only with the Christoffel symbols.

Now let us implement the additional point (1) mentioned in the Introduction, as far as $\Delta_D^{(3)}$ is concerned. It is easy to show that $\Delta_D^{(3)}$ can be canceled by a local counterterm. Indeed we remark that it is mapped into zero by the Bardeen–Zumino map, which maps a-cocycles of Σ_D into a-cocycles of Σ_L (see Ref. 5). Indeed let us write Eq. (5.40) as

$$\Delta_D^{(3)} = \int \mathrm{Tr}(d\Lambda\, G^{(3)}(\Gamma)) \tag{6.2}$$

and let us consider the functional

$$S^{(3)} = \int_0^1 dt \int_x \mathrm{Tr}(HG^{(3)}(\Gamma_t)), \tag{6.3}$$

where $\Gamma_t = e^{-tH}\Gamma e^{tH} + e^{-tH} de^{tH}$ and H is the logarithm of the vielbein matrix. We obtain

$$\Sigma_D S^{(3)} = \Delta_D^{(3)}, \quad \Sigma_L S^{(3)} = 0. \tag{6.4}$$

In fact, the cocycle (5.40) is a sort of "fossil" of the prehistory of the Lorentz bundle, upon which a gravitation theory is constructed. A Lorentz bundle is a reduced subbundle of the linear frame bundle, where the "gauge" group is GL(4). From this point of view the a-cocycles $\Delta_D^{(i)}$ ($i = 1,2,3$) are understandable.[17]

Finally let us quote the following coboundary:

$$\hat{\Delta}_D = \int \mathrm{Tr}(d\Lambda\, d\Gamma) A. \tag{6.5}$$

Here A is a one-form $A_m\, dx^m$, where A_m is any vector field with weight 0. We shall elaborate on it as a useful illustration of Theorem 3.1. It is a coboundary since

$$\begin{aligned}\hat{\Delta}_D &= \int_x \mathrm{Tr}(d\Lambda(R + \Gamma\Gamma))A \\ &= \Sigma_D \int_x \mathrm{Tr}\Big(\Big(R + \frac{1}{3}\Gamma\Gamma\Big)\Gamma\Big) A,\end{aligned} \tag{6.6}$$

where R is the curvature two-form. If $A_m = T^l_{ml}$, $\hat{\Delta}_D$ is a coboundary we may disregard. But if A_m is an Abelian gauge field, the local action on the right-hand side of Eq. (6.6) is not gauge invariant: removing the coboundary $\hat{\Delta}_D$ interferes with gauge invariance. This is an indication of the existence of an a-cocycle of the coupled operation $\Sigma_D + \Sigma_G$. We shall clarify this point in the next section.

VII. THE COHOMOLOGY OF $\Sigma_D + \Sigma_L + \Sigma_G$

We are now able to find all the a-cocycles of the whole cohomological system $\Sigma_D + \Sigma_L + \Sigma_G$. Let us consider first $\Sigma_D + \Sigma_L$. On the basis of Theorem 3.1, we know that the a-cocycles of $\Sigma_D + \Sigma_L$ are determined by the a-cocycles of Σ_D that admit a Lorentz partner and by the D-a-cocycles of Σ_L. The latter are well known because they correspond to the Lorentz anomalies, which are computed by understanding D-invariance.[3–5] In four dimensions there exist no D-a-cocycles of Σ_L (see, however, the comment on the mixed anomaly at the end of this section).

As for the a-cocycles of Σ_D, those belonging to the second family are certainly admissible because they are Σ_L-invariant. The first family requires a closer examination. Let us write the generic a-cocycle belonging to it as

$$\check{\Delta}_D = \int \partial_l \xi^l \sqrt{g}\mathscr{L}. \tag{7.1}$$

We split the family into three sets.

(1st set) $\check{\Delta}_D^{(i)}$, $i = 1,2,\ldots$:

$$\Sigma_L\, {}_1\check{\Delta}_D^{(i)} = 0, \quad \Sigma_G\, {}_1\check{\Delta}_D^{(i)} = 0.$$

(2nd set) ${}_2\check{\Delta}_D^{(j)}$, $j = 1,2,\ldots$: $\Sigma_L\, {}_2\check{\Delta}_D^{(i)} \neq 0$.

Examples:

$$\mathscr{L} = \mathrm{Tr}(\omega_m \omega_n \omega_l \omega_r) g^{mn} g^{lr}$$

or

$$\mathrm{Tr}(\omega_m \omega_n) g^{mn}\, \mathrm{Tr}(V_l V_r) g^{lr},$$

where $V_l = V_l^a T^a$ is a gauge field and $\omega = \omega^{ab}\Sigma_{ab}$ is the Lorentz connection.

(3rd set) ${}_3\check{\Delta}_D^{(k)}$, $k = 1,2,\ldots$:

$$\Sigma_L\, {}_3\check{\Delta}_k^{(i)} = 0, \quad \Sigma_G\, {}_3\check{\Delta}_k^{(i)} \neq 0.$$

Example: $\mathscr{L} = \mathrm{Tr}(V_m V_n V_l V_r) g^{mn} g^{lr}$.

Let us consider, for example,

$$\begin{aligned}\Sigma_L\, {}_2\check{\Delta}_D^{(i)} &= \Sigma_L \int_x \partial_l \xi^l \sqrt{g} g^{mn} g^{lr}\, \mathrm{Tr}(\omega_m \omega_n \omega_l \omega_r) \\ &= -4 \int_x \partial_l \xi^l \sqrt{g} g^{mn} g^{lr}\, \mathrm{Tr}(\partial_m u\, \omega_n \omega_l \omega_r).\end{aligned}$$

There exists no P-functional ${}_2\check{\Delta}_L^{(1)}$ linear in u_a^b such that $\Sigma_D\, {}_2\check{\Delta}_L^{(1)} + \Sigma_L\, {}_2\check{\Delta}_D^{(1)} = 0$. Therefore ${}_2\check{\Delta}_D^{(1)}$ is not an admissible a-cocycle of Σ_D w.r.t. $\Sigma_D + \Sigma_L$. One easily realizes that the same is true for all the a-cocycles of the second set, while those of the first and third set are all admissible. Therefore we have completely determined the a-cocycles of the operator $\Sigma_D + \Sigma_L$.

When also Σ_G is taken into account, we again apply Theorem 3.1, where the role of Σ_S is now played by $\Sigma_D + \Sigma_L$ and that of Σ_R by Σ_G. Using the same arguments as above we easily find that the third set must be excluded, too, while we must add the diffeomorphisms and Lorentz-invariance-preserving a-cocycles of Σ_G, i.e., the usual and well-known gauge anomalies.[1,18] Finally the nontrivial cohomology classes of $\Sigma_D + \Sigma_L + \Sigma_G$ are determined by the a-cocycles of Σ_D belonging to the second family and to the first set of the first family, and by the usual gauge anomalies: the expected or, perhaps, the desired result.

As for the a-cocycles of the first set of the first family a comment is in order. They can be expressed as

$$ {}_1\check{\Delta}_D^{(i)} = \Sigma_D \check{C}^{(i)}, \quad i = 1,2,\dots, \tag{7.2}$$

where

$$\check{C}^{(i)} = \int \sqrt{g} \ln \sqrt{g} \mathscr{L}^{(i)}, \quad i = 1,2,\dots . \tag{7.3}$$

Technically speaking, according to our definition, the $\Delta_D^{(i)}$'s are a-cocycles since $C^{(i)}$'s in Eq. (7.3) are local but nonpolynomial functionals. However, Eq. (7.2) tells us that there are regularizations free from such anomalies. Nevertheless, the $\Delta_D^{(i)}$ are not devoid of physical interest since, as shown in Ref. 6, they are strictly connected with Weyl anomalies.

Above we did not mention the so-called mixed Lorentz anomaly

$$\check{\Delta}_L = \int \mathrm{Tr}(du\, d\omega) A = \Sigma_L \int \mathrm{Tr}\Big(\omega R - \frac{1}{3}\omega\omega\omega\Big) A, \tag{7.4}$$

where $u = u^{ab}\Sigma_{ab}$, because it is a coboundary exactly in the same way as $\hat{\Delta}_D$ [Eq. (6.5)] is. We recall that we have shown in Ref. 6 that they are mapped into each other by the Bardeen–Zumino map. Now, let us recall Eq. (6.6) and observe that

$$\Sigma_G \int \mathrm{Tr}\Big(\omega R \frac{1}{3}\omega\omega\omega\Big) A$$
$$= \Sigma_G \int \mathrm{Tr}\Big(\Gamma\Big(R + \frac{1}{3}\Gamma\Gamma\Big)\Big) A = \int \lambda\, \mathrm{Tr}(RR) \equiv \hat{\Delta}_G,$$

since $\Sigma_G A = d\lambda$. Therefore this a-cocycles of $\Sigma_D + \Sigma_L + \Sigma_G$ may appear either as a gauge anomaly or as a Lorentz anomaly or as a D-anomaly. We may reduce it to the form we wish through a simple redefinition of the vertex generating functional. Observe that Eq. (7.5) implies that the Bardeen–Zumino functional necessary in order to map $\hat{\Delta}_L$ and $\hat{\Delta}_G$ into each other,[5,6] is a P-functional. Finally we remark that in the above list of a-cocycles of $\Sigma_D + \Sigma_L + \Sigma_G$, this a-cocycle appears in the form $\hat{\Delta}_G$ simply because of the order among Σ_D, Σ_L, and Σ_G we have chosen in applying Theorem (3.1).[19] Needless to say, this order is arbitrary.

APPENDIX A: TRANSLATIONS

Rigid translations are the Abelian subgroup τ of D obtained by restricting the parameters $\xi^m(x)$ to constant values b^m. We can construct the cohomology operator Σ_T in the usual way. Since the group is Abelian the "ghosts" b^m do not transform. If the differential space is the space of local P-functionals, there are a-cocycles of Σ_T. For, if $\mathfrak{a}_m$ is any local expression not expressible as a total derivative, then

$$\Sigma_T \int b^m \mathfrak{a}_m = 0, \quad \int b^m \mathfrak{a}_m \neq \Sigma_T C, \quad \forall \text{ local } C. \tag{A1}$$

We can avoid these anomalies by enlarging the differential space to include all integrated polynomial expressions of the fields, their derivatives and the coordinates x^m. Then it is easy to see that

$$\int b^m \mathfrak{a}_m = -\Sigma_T \int x^m \mathfrak{a}_m. \tag{A2}$$

Of course in this way we get more cocycles of Σ_T but it turns out that they are all coboundaries. Indeed let

$$\mathfrak{a}_m = C_{ml_1\cdots l_n} x^{l_1}\cdots x^{l_n}, \tag{A3}$$

where $C_{ml_1l_2\cdots l_n}$ is any general local expression (it may have other saturated or unsaturated indices besides those written down), symmetric in $l_1\cdots l_n$. The consistency condition can be written as

$$\Sigma_T \int_x b^m C_{ml_1\cdots l_n} x^{l_1}\cdots x^{l_n} = -\int b^m b^p \partial_p C_{ml_1\cdots l_n} x^{l_1}\cdots x^{l_n}$$
$$= n \int b^m b^p C_{mpl_2\cdots l_n} x^{l_2}\cdots x^{l_n}. \tag{A4}$$

Therefore either (1) $C_{ml_1\cdots l_n}$ is totally symmetric in $m, l_1,\dots,l_n$, or (2) $C_{mpl_2\cdots l_n} x^{l_2}\cdots x^{l_n}$ is a derivative.

Case (1):

$$\int C_{ml_1\cdots l_n} x^m x^{l_1}\cdots x^{l_n} \neq 0$$

and

$$\Sigma_T \int C_{ml_1\cdots l_n} x^{l_1}\cdots x^{l_n} x^m = (n+1) \int b^m C_{ml_1\cdots l_n} x^{l_1}\cdots x^{l_n}, \tag{A5}$$

which shows that the cocycle is a coboundary.

Case (2): $C_{ml_1\cdots l_n}$ is an nth-order derivative and, integrating by parts repeatedly in the initial expression, we are reduced to Eqs. (A1) and (A2).

The proof can be easily extended to expressions containing finite sums of terms of the type (A3). Therefore all cocycles of Σ_T are coboundaries.

Now let us define the coupled cohomology of Σ_T and Σ_A ($\bar{\Sigma}_A$). If we set

$$\Sigma_T \alpha_l{}^p = 0, \quad \Sigma_A b^m = b^p \alpha_p{}^m, \quad \bar{\Sigma}_A b^m = b^p \alpha_p{}^m, \tag{A6}$$

we get

$$(\Sigma_T + \Sigma_A)^2 = 0 \quad \text{and} \quad (\Sigma_T + \bar{\Sigma}_A)^2 = 0. \tag{A7}$$

For D we can do the same, provided that we define

$$\Sigma_T \xi^l = b^m \partial_m \xi^l, \quad \Sigma_D b^m = 0. \tag{A8}$$

Then

$$(\Sigma_T + \Sigma_D)^2 = 0. \tag{A9}$$

By applying Corollary 3.2, we can now conclude that the admissible a-cocycles of Σ_A, $\bar{\Sigma}_A$, and Σ_D lie in the restricted differential space of the local P-functionals.

APPENDIX B: THE COHOMOLOGY OF Σ_A

${}^w\mathfrak{a}$ in Eq. (4.1) is a local polynomial whose transformation law under Σ_A is the following:

$$\Sigma_A\, {}^w\mathfrak{a}^{mk_1\cdots k_r}_{nl_1\cdots l_s} = x^p\alpha_p{}^q\,\partial_q\, {}^w\mathfrak{a}^{mk_1\cdots k_r}_{nl_1\cdots l_s} + \alpha_n{}^p\, {}^w\mathfrak{a}^{mk_1\cdots k_r}_{pl_1\cdots l_s}$$

$$- \alpha_l{}^m\, {}^w\mathfrak{a}^{lk_1\cdots k_r}_{nl_1\cdots l_s} + \sum_{i=1}^{s} \alpha_{l_i}{}^p\, {}^w\mathfrak{a}^{mk_1\cdots k_r}_{nl_1\cdots p\cdots l_s}$$

$$- \sum_{j=1}^{r} \alpha_p{}^{k_j}\, {}^w\mathfrak{a}^{mk_1\cdots p\cdots k_r}_{nl_1\cdots l_s} + w\alpha_p{}^p\, {}^w\mathfrak{a}^{mk_1\cdots k_r}_{nl_1\cdots l_s}. \tag{B1}$$

Using this equation and Eq. (2.14) one can write down explicitly the consistency condition $\Sigma_A \Delta_A = 0$. Integrating by parts and differentiating twice w.r.t. α^n_m and α^l_k, one gets

$$0 = \sum_{w,r,s} \sum_{\substack{k_1\cdots k_2 \\ l_1\cdots l_s}} C^{\,l_1\cdots l_s}_{\,k_1\cdots k_r} \int \Big[\delta^k_n\, {}^w\mathfrak{a}^{mk_1\cdots k_r}_{ll_1\cdots l_s} - \delta^m_l\, {}^w\mathfrak{a}^{kk_1\cdots k_r}_{nl_1\cdots l_s}$$

$$+ \sum_{i=1}^{s} \big(\delta^k_{l_i}\, {}^w\mathfrak{a}^{mk_1\cdots k_r}_{nl_1\cdots l\cdots l_s} - \delta^m_{l_i}\, {}^w\mathfrak{a}^{kk_1\cdots k_r}_{ll_1\cdots n\cdots l_s}\big)$$

$$+ \sum_{j=1}^{r} \big(\delta^{k_j}_l\, {}^w\mathfrak{a}^{mk_1\cdots k\cdots k_r}_{nl_1\cdots l_s} - \delta^{k_j}_n\, {}^w\mathfrak{a}^{kk_1\cdots m\cdots k_r}_{ll_1\cdots l_s}\big)$$

$$+ (w-1)\big(\delta^k_l\, {}^w\mathfrak{a}^{mk_1\cdots k_r}_{nl_1\cdots l_s} - \delta^m_n\, {}^w\mathfrak{a}^{kk_1\cdots k_r}_{ll_1\cdots l_s}\big)\Big]. \tag{B2}$$

Let us denote by $\mathfrak{b}^{kmk_1\cdots k_r}_{nll_1\cdots l_s}$ the object contained inside the square brackets.

We are going to show now that the only solution of this equation is $\mathfrak{b} = 0$ identically. To see this we perform on the field a generic finite transformation $a \in D_A$. We get

$$\sum_{w,r,s} \sum_{\substack{k_1\cdots k_r \\ l1\cdots l_s}} \int C^{\,l_1\cdots l_s}_{\,k_1\cdots k_r}\, (\det a)^{w-1} a_{l_1}^{\,l_1'} \cdots a_{l_s}^{\,l_s'}$$

$$\times (a^{-1})^{k_1}_{k_1'} \cdots (a^{-1})^{k_r}_{k_r'}\, {}^w\mathfrak{b}^{kmk_1'\cdots k_r'}_{nll_1'\cdots l_s'}. \tag{B3}$$

This is the same as Eq. (B2) except that the C coefficients have been transformed. Since $C^{\,l_1\cdots l_s}_{\,k_1\cdots k_r}$ does not contain (constant) tensors (we have excluded it from the beginning, in Sec. IV), we can conclude that terms with different r, s, and w must vanish separately. Likewise we can conclude that Eq. (4.2) implies

$${}^w\mathfrak{b}^{kmk_1\cdots k_r}_{nll_1\cdots l_s} = 0, \quad \forall w,r,s. \tag{B4}$$

The possibility that B be a global derivative corresponds to a vanishing cocycle. From the form of ${}^w\mathfrak{b}$ one sees that the most general ${}^w\mathfrak{a}$ satisfying Eq. (B4) can be written as

$${}^w\mathfrak{a}^{kk_1\cdots k_r}_{ll_1\cdots l_s} = \delta^k_l \mathring{\mathfrak{a}}^{k_1\cdots k_r}_{l_1\cdots l_s} + \sum_{a=1}^{r} \delta^{k_a}_l \overset{a}{\mathfrak{a}}{}^{k_1\cdots k\cdots k_r}_{l_1\cdots l_s} + \sum_{b=1}^{s} \delta^k_{l_b} \underset{b}{\mathfrak{a}}{}^{k_1\cdots k_r}_{l_1\cdots l\cdots l_s}, \tag{B5}$$

where $\mathring{\mathfrak{a}}$, $\overset{a}{\mathfrak{a}}$, $\underset{b}{\mathfrak{a}}$ are local polynomials of the fields and their derivative to be determined. Substituting (B5) into Eqs. (B4) and (B2) we get a sum of terms involving these unknown polynomials multiplied by two Kronecker δ's. It can vanish only if (1) all the coefficients of distinct products of δ's vanish and (2) $\mathring{\mathfrak{a}}$, $\overset{a}{\mathfrak{a}}$ and $\underset{b}{\mathfrak{a}}$ further factorize into δ factors.

Case (1): The independent equations that one derives are

$$\mathring{\mathfrak{a}}^{k_1\cdots k_r}_{l_1\cdots l_s} = -\overset{a}{\mathfrak{a}}{}^{k_1\cdots k_r}_{l_1\cdots l_s}(w-1), \quad \forall a,$$

$$\mathring{\mathfrak{a}}^{k_1\cdots k_r}_{l_1\cdots l_s} = \underset{b}{\mathfrak{a}}{}^{k_1\cdots k_r}_{l_1\cdots l_s}(w-1), \quad \forall b, \tag{B6}$$

$$\overset{a}{\mathfrak{a}}{}^{k_1\cdots k_r}_{l_1\cdots l_s} = -\underset{b}{\mathfrak{a}}{}^{k_1\cdots k_r}_{l_1\cdots l_s}, \quad \forall a,b.$$

Equation (B5) becomes

$${}^w\mathfrak{a}^{kk_1\cdots k_r}_{ll_1\cdots l_s} = (w-1)\delta^k_l \bar{\mathfrak{a}}^{k_1\cdots k_r}_{l_1\cdots l_s} - \sum_{a=1}^{r} \delta^{k_a}_l \bar{\mathfrak{a}}^{k_1\cdots k\cdots k_r}_{l_1\cdots l_s} + \sum_{b=1}^{s} \delta^k_{l_b} \bar{\mathfrak{a}}^{k_1\cdots k_r}_{l_1\cdots l_s}. \tag{B7}$$

But

$$\Sigma_A \int \bar{\mathfrak{a}}^{k_1\cdots k_r}_{l_1\cdots l_s} = \int \alpha^l_k \Big((w-1)\delta^k_l \bar{\mathfrak{a}}^{k_1\cdots k_r}_{l_1\cdots l_s} + \sum_{i=1}^{r} \delta^r_{l_i}\, \bar{\mathfrak{a}}^{k_1\cdots k_r}_{l_1\cdots l\cdots l_s}$$

$$- \sum_{j=1}^{r} \delta^{k_j}_l \bar{\mathfrak{a}}^{k_1\cdots k\cdots k_r}_{l_1\cdots l_s}\Big). \tag{B8}$$

Therefore all the solutions of the consistency condition (B2) pertaining to case (1) are coboundaries. When $r = s = 0$, the above treatment is meaningless. This particular case is treated in Sec. IV.

Case (2): No new solutions from this case.

Therefore the nontrivial cohomology classes of Σ_A are determined by Eq. (4.2).

APPENDIX C: PROOF OF THE SUFFICIENCY CONDITION

In this Appendix we prove the sufficiency condition in part I of Theorem 5.1. To this end we recall Eq. (5.1) and all the specifications made at the beginning of Sec. V. To Eq. (5.1) we add all possible terms coming from dP^1_3, for a class A P^1_3. The density of $Q^1_4 - dP^1_3$ is then

$$\widetilde{Q} = \partial\partial\xi(B_1 + \Gamma C_1 + \Gamma\Gamma D_1 + \partial\Gamma E_1 + \Gamma\Gamma\Gamma F_1 + \partial\Gamma\Gamma G_1$$

$$+ \partial\partial\Gamma H_1) + \partial\partial\partial\xi(B_2 + \Gamma C_2 + \Gamma\Gamma D_2 + \partial\Gamma E_2)$$

$$+ \partial\partial\partial\partial\xi(B_3 + \Gamma C_3) + \partial\partial\partial\partial\partial\xi B_4, \tag{C1}$$

where all B_i, C_i, etc. are D-covariant tensors with D-weight 1.

Now, the relevant consistency condition is

$$\hat{\Sigma}_D \widetilde{Q} = 0. \tag{C2}$$

This means, in concise form,

$$\partial\partial\xi\partial\partial\xi X + \partial\partial\xi\partial\partial\partial\xi Y + \partial\partial\xi\partial\partial\partial\partial\xi Z + \partial\partial\partial\xi\partial\partial\partial\xi W = 0, \tag{C3}$$

where X, Y, Z, and W depend on Γ, $\partial\Gamma$ and on the covariant coefficients. It is clear that in order for Eq. (C3) to be satisfied, X and W must either vanish or have suitable symmetry properties in the indices. After inspecting their explicit form one can conclude that Y and Z can only vanish. Briefly

$$X \overset{s}{=} 0, \quad Y = 0, \quad Z = 0, \quad W \overset{s}{=} 0, \tag{C4}$$

where $\overset{s}{=}$ means equal up to symmetry properties of the indices. The first two equations of (C4) break down in turn into more independent equations according to the powers of Γ and $\partial\Gamma$. We are going to find all the solutions of these equations and show that they are all coboundaries.

To this end it is very useful to write down all the coboundaries of $\hat{\Sigma}_D$ belonging to the second family. They can be written as follows:

$$\begin{aligned}
&\mathscr{C}_1 = \Gamma M_1, \\
&\mathscr{C}_2 = \Gamma\Gamma N_1, \quad \mathscr{C}_5 = \partial\Gamma N_2, \\
&\mathscr{C}_3 = \Gamma\Gamma\Gamma R_1, \quad \mathscr{C}_6 = \partial\Gamma\Gamma R_2, \quad \mathscr{C}_8 = \partial\partial\Gamma R_3, \quad \mathscr{C}_{10} = \partial\partial\Gamma\Gamma S_5, \\
&\mathscr{C}_4 = \Gamma\Gamma\Gamma\Gamma S_1, \quad \mathscr{C}_7 = \partial\Gamma\Gamma\Gamma S_2, \quad \mathscr{C}_9 = \partial\Gamma\partial\Gamma S_4, \quad \mathscr{C}_{11} = \partial\partial\partial\Gamma S_6,
\end{aligned} \tag{C5}$$

where M_1, N_i, R_i, S_i are covariant tensors of D-weight 1 and dimensions 3, 2, 1, and 0, respectively.

Our strategy consists in identifying all the solutions of Eqs. (C4) corresponding to the coboundaries (C5) and proving that they are the only ones. For example, if we set all the coefficients equal to zero except G_1 and E_2, we have a solution of Eqs. (C4),

$$G_{1m}{}^{p_1p_2n_1}{}_{l}{}^{n_2s_1s_2q}{}_{r} = 2\delta^{n_1}_{m} E_2{}^{s_1s_2q}_{r}{}^{p_1p_2n_2}_{l} - \delta^{p_1}_{l} E_2{}^{s_1s_2q}_{r}{}^{n_1n_2p_2}_{m}, \tag{C6}$$

where $E_{2r}{}^{s_1s_2qp_1p_2n_2}_{l} = E_{2l}{}^{p_1p_2n_2s_1s_2q}_{r}$. This solution of Eq. (C2) corresponds to a coboundary, precisely to $\hat{\Sigma}_D \mathscr{C}_9$, provided that we identify E_2 with S_4. Similarly we find other solutions of Eq. (C2) corresponding to

$$\begin{array}{llll}
\hat{\Sigma}_D \mathscr{C}_1 & \text{if} \;\; B_1 = M_1, & \hat{\Sigma}_D \mathscr{C}_6 & \text{if} \;\; C_2 = E_1 = R_2 \\
\hat{\Sigma}_D \mathscr{C}_2 & \text{if} \;\; C_1 = 2N_1, & \hat{\Sigma}_D \mathscr{C}_7 & \text{if} \;\; D_2 = \tfrac{1}{2} G_1 = S_2, \\
\hat{\Sigma}_D \mathscr{C}_3 & \text{if} \;\; D_1 = 3R_1, & \hat{\Sigma}_D \mathscr{C}_8 & \text{if} \;\; B_3 = R_3, \\
\hat{\Sigma}_D \mathscr{C}_4 & \text{if} \;\; F_1 = 4S_1, & \hat{\Sigma}_D \mathscr{C}_{10} & \text{if} \;\; C_3 = H_1 = S_5, \\
\hat{\Sigma}_D \mathscr{C}_5 & \text{if} \;\; B_2 = N_2, & \hat{\Sigma}_D \mathscr{C}_{11} & \text{if} \;\; B_4 = S_6,
\end{array} \tag{C7}$$

All these solutions must be contained in the most general solution of Eq. (C2) [or (C3) or (C4)]. Since they can be multiplied by arbitrary numerical coefficients we can subtract them from the latter, which, in this way, looses any dependence on the covariant coefficients except E_1, G_1, and H_1. When one considers again Eq. (C2) with only these coefficients surviving, one easily sees that the only solution is $E_1 = G_1 = H_1 = 0$. Therefore the only solutions of Eq. (C2) are the coboundaries (C5).

APPENDIX D: SPECIFICATION OF THE DERIVATION OF (5.27)

The derivation of formula (5.27) deserves a specification. $\hat{\Sigma}_D dP^2_2$ in Eq. (5.25) can be written as follows, in terms of dual tensors:

$$\hat{\Sigma}_D \partial_l (\xi^i \xi^j A_{ij}{}^{lr} + \xi^i \partial_p \xi^j B_{ij}{}^{plr} + \xi^i \partial_{p_1} \partial_{p_2} \xi^j C_{ij}^{p_1p_2lr} + \cdots). \tag{D1}$$

Here we have written down explicitly only the first terms [see Eq. (5.23)]. Equation (5.25) implies $\partial_l A_{ij}{}^{lr} = 0$, consequently A can be absorbed into B. In general, Eq. (5.25) implies also $\partial_l B_{ij}{}^{plr} = 0$. That is not the case when $B_{ij}^{plr} = \delta^p_i \hat{B}_j{}^{lr}$, because $\hat{\Sigma}_D (\xi^i \partial_i \xi^j) = 0$. Moreover we must have

$$\delta^{p_1}_i \delta^{p_2}_l \hat{B}_j{}^{lr} + \partial_l C_{ij}{}^{p_1p_2lr} = 0.$$

Either $B = C = 0$ or

$$\hat{B}_j{}^{lr} = \partial_p \hat{C}_j{}^{plr}, \tag{D2}$$

$${}^iQ^1_4 = \hat{\Sigma}_D \mathscr{C}_i, \quad i = 1,\ldots,11,$$

where

where $\hat{C}$ is completely antisymmetric in p, l, r. Now we can apply the usual argument and, after repeated applications, obtain Eq. (5.27).

The cohomology defined by Eq. (5.32) can be solved along the same lines as the main problem. With a theorem analogous to 5.1 we prove that P^1_2 is a coboundary if and only if it is a solution of $\hat{\Sigma}_D (P^1_2 - dR^1_1) = 0$ for some class A R^1_1 (we recall that P^1_2 is class A). The nontrivial solutions are contained in the set of P^1_2's such that $P^2_1 \neq 0$. Since dP^2_1 must be class A we have a theorem analogous to (5.5): the dual tensor of dP^2_1 can be written as

$$\partial_l (\partial_q \xi^i \partial_{p_1} \partial_{p_2} \xi^j) F_{ij}{}^{qp_1p_2lmn},$$

which implies that F is a constant tensor that can be written $F_{ij}^{qp_1p_2lmn} = \epsilon^{p_1lmn} \hat{F}_{ij}^{qp_2}$, and $\hat{F}_{ij}^{qp_2}$ and is made of Kronecker δ's. Therefore up the total differentials we have only two possibilities:

$$P^{2(1)}_1 = \mathrm{Tr}(\Lambda)\mathrm{Tr}(d\Lambda) \quad \text{and} \quad P^{2(2)}_1 = \mathrm{Tr}(\Lambda\, d\Lambda). \tag{D3}$$

They correspond to two distinct a-cocycles of Eq.(5.34).

[1] R. Stora, Cargèse Lectures, 1983; B. Zumino, Les Houches lectures, 1983.
[2] L. Alvarez-Gaumé and E. Witten, Nucl. Phys. B **234**, 269 (1984).
[3] F. Langouche, T. Schücker, and R. Stora, Phys. Lett. B **145**, 342 (1984).
[4] L. Baulieu and J. Thierry-Mieg, Phys. Lett. B **145**, 53 (1984).
[5] W. A. Bardeen and B. Zumino, Nucl. Phys. B **244**, 421 (1984).
[6] L. Bonora, P. Pasti, and M. Tonin, Phys. Lett. B **149**, 346 (1984).
[7] J. Wess and B. Zumino, Phys. Lett. B **37**, 95 (1971).
[8] C. Becchi, A. Rouet, and R. Stora, Ann. Phys. (NY) **98**, 287 (1976).
[9] L. Bonora and P. Cotta-Ramusino, Commun. Math. Phys. **87**, 589 (1983).
[10] L. D. Faddeev, Phys. Lett. B **145**, 81 (1984).
[11] G. Bandelloni, Nuovo Cimento A **88**, 1 (1985).
[12] W. Greub, S. Halperin, and R. Vanstone, *Connections, Curvature and Cohomology* (Academic, New York, 1976).

[13] The summation convention is defined by $(\Gamma\Gamma)^l_m = \Gamma^k_m \Gamma^l_k$. The context will make clear whether Γ is the matrix one-form or the symbol Γ^l_{mn} with all the indices dropped.
[14] We refer to the duality defined through the Hodge star operator.
[15] These arguments are exact only if we disregard the relation among connection and metric, as explained at the beginning of the section. This is the only place where we treat connection and metric as uncorrelated fields (see also the remark at the end of Sec. V).

[16] C. Becchi, A. Rouet, and R. Stora, *Field Theory Quantization and Statistical Physics*, edited by E. Tirapequi (Reidel, Dordrecht, 1981).
[17] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley–Interscience, New York, 1963), Vol. I, p. 83.
[18] J. Thierry-Mieg, Phys. Lett. B **147**, 430 (1984).

[19] In the form $\hat{\Delta}_G$ the Weyl partner vanishes; the same is not true for $\hat{\Delta}_L$ and $\hat{\Delta}_D$ (see Ref. 8).

A path-integral–Riemann-space approach to the electromagnetic wedge diffraction problem

Richard W. Ziolkowski
Electronics Engineering Department, Lawrence Livermore National Laboratory, P. O. Box 808, L-156, Livermore, California 94550

(Received 24 October 1984; accepted for publication 30 April 1986)

A path integral constructed over a particular Riemann space is developed and applied to two-dimensional wedge problems. This path-integral–Riemann-space (PIRS) approach recovers the exact solutions of the heat conduction and the corresponding electromagnetic wedge problems. A high-frequency asymptotic evaluation of the PIRS electromagnetic wedge solution returns the standard geometrical theory of diffraction (GTD) results. Ramifications of this approach and its relationships with known path-integral methods are examined.

I. INTRODUCTION

Many quantum mechanical applications of path integrals defined on multiconnected spaces have appeared in the literature.[1–5] Similarly, using a double-sheeted Riemann surface Buslaev[6] established the viability of the path-integral approach to the scattering of electromagnetic waves from smooth conductors. However, in spite of the known importance of the multiconnected space description of diffraction phenomena (see Sommerfeld[7] or Carslaw[8]), the application of an analogous path-integral approach to electromagnetic diffraction problems has been neglected. It is the object of this paper to demonstrate the utility of a path-integral–Riemann-space approach in wedge diffraction problems and to point out several interesting aspects of the resultant representations of the solutions.

In Secs. II–V, a path-integral–Riemann-space (PIRS) approach is developed and applied to the electromagnetic diffracting (perfectly conducting) wedge problem. As in Buslaev[6] and Lee,[9] the diffraction problem is first transformed to its equivalent heat conduction problem. The latter is treated with the PIRS approach. The transform of the resultant expression returns the exact wedge diffraction solution. A high-frequency asymptotic approximation of the PIRS solution is given in Sec. VI. It recovers the results given by Keller's geometrical theory of diffraction (GTD).[10] In Sec. VII, several properties of the PIRS solution to the electromagnetic and heat conduction wedge problems are described. Relations based upon the multivaluedness of the solutions are derived that demonstrate that the modification of free-space by the wedge leads to the diffraction effects. Moreover, it is shown that the half-plane propagator satisfies a transition condition that is characteristic of the underlying Riemann space and is associated with a particular Riemann–Hilbert problem.[11] The relationships of the PIRS approach with analogous quantum mechanical methods are also discussed. For instance, the connection between the PIRS method and the constrained path-integral approach[12–16] is established. It indicates that a PIRS wedge analysis may prove useful for studies of fractional charge quantization. Other salient features of the PIRS approach suggest its applicability to related problems of interest involving entangled polymers in molecular biology, Ising models in statistical mechanics, and soliton and instanton models in quantum field theory.

II. WEDGE DIFFRACTION PROBLEM

A. Problem configuration

Consider in two dimensions the diffraction of the field due to a unit point source by a perfectly conducting wedge with exterior angle $\beta\pi$, $1<\beta<2$. The electric field vector is assumed to be parallel to the edge of the wedge (E-polarized field). This is equivalent to the three-dimensional diffraction problem in which a line source is parallel to the edge of a wedge of infinite extent. The scattered field is also E-polarized and is assumed to satisfy the radiation condition at infinity.

A polar coordinate system is erected whose origin is located at the edge of the wedge. Angles measured in a counterclockwise direction from the upper edge of the wedge defined to be $\theta=0$ are positive. The lower edge is defined by $\theta=\beta\pi$. The physical space $[0,\infty[\times[0,\beta\pi]$, exterior to the wedge, is denoted by P. The observation point is located at $\mathbf{r}=(r,\theta)$; the unit source $\delta(\mathbf{r}-\mathbf{r}_0)$ is located at $\mathbf{r}_0=(r_0,\theta_0)$. This geometry is shown in Fig. 1.

B. The Riemann spaces P_2 and P_∞

The original diffraction problem in the physical space P is simplified by considering diffraction in a space P_2 con-

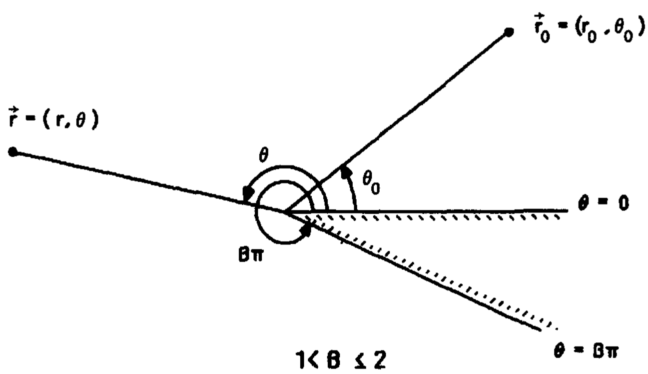


FIG. 1. Geometry of the two-dimensional diffraction by a wedge problem.

 0022-2488/86/092271-11$02.50

structed as follows. Take two replicas of P, say P_+ and P_-, and join them along the boundary Σ of the wedge. Then $P_2 = P_+ \cup P_- \cup \Sigma$. The spaces P_+ and P_- will be called, respectively, the upper and lower sheets of P_2; the space P_+ is identified with the physical space P. To suggest pictorally the two sheets, the "edge" of P_- is drawn outside of that of P_+ as illustrated in Fig. 2(a). A function $U(\mathbf{r})$ over P_2 will be a wave function if it satisfies the Helmholtz equation

$$\{\Delta + k^2\}U(\mathbf{r}) = 0 \tag{2.1}$$

over P_+ and P_- (open sets) and if the limiting values U_+ and U_- of $U(\mathbf{r})$, when $\mathbf{r}$ approaches Σ from P_+ and P_-, are opposite and if the corresponding normal derivatives on Σ toward P_+ and P_- are continuous:

$$\begin{matrix}\text{boundary}\\ \text{conditions}\end{matrix}\begin{cases} U_+ + U_- = 0, \\ \partial_n U_+ = \partial_n U_-, \end{cases} \text{ on } \Sigma . \tag{2.2}$$

The space P_2 is a Riemann surface, and its use here is very similar to the device introduced by Sommerfeld[7] for the half-plane problem and by Buslaev[6] for the convex body case. Natural coordinates in P_2 are the distance r to the origin and the polar angle θ counted from the upper edge of the wedge. This angle varies from 0 to $\Omega = 2\pi\beta$ and the angle $\theta = 2\pi\beta$ is identified with $\theta = 0$. This geometry is shown in Fig. 2(b).

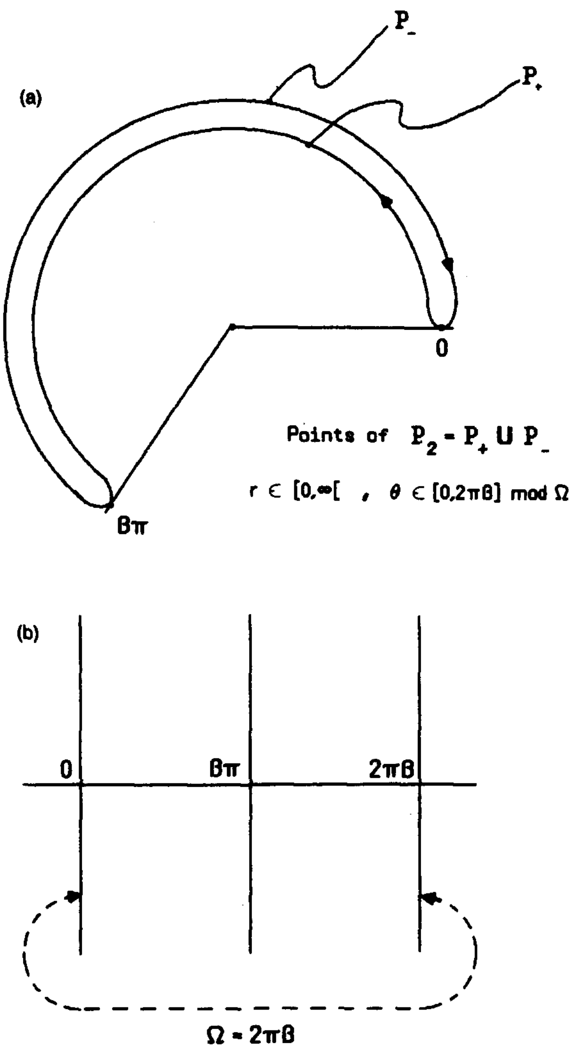


FIG. 2. Representations of the space P_2: (a) as a two-sheeted Riemann surface and (b) as its angular extent.

Note that the angles $\theta = 0, \beta\pi$ have no special properties. In fact, the effect of introducing P_2 may be considered as "erasing" the boundaries of the wedge. The boundary conditions are satisfied by locating an image source on P_2 (specifically, on the lower sheet P_-) at $\mathbf{r}_0' = (r_0, -\theta_0)$. The desired field can then be decomposed as

$$U(\mathbf{r}) = K(\mathbf{r},\mathbf{r}_0) - K(\mathbf{r},\mathbf{r}_0') , \tag{2.3}$$

where, for example, $K(\mathbf{r},\mathbf{r}_0)$ represents the field at $\mathbf{r}$ due to the source point at $\mathbf{r}_0$. Since reciprocity must be satisfied, $K(\mathbf{r},\mathbf{r}_0')$ must be of the same form as $K(\mathbf{r},\mathbf{r}_0)$; the former is obtained from the latter by a simple substitution $\theta_0 \to -\theta_0$. Hence, it will only be necessary to consider the function $K(\mathbf{r},\mathbf{r}_0)$.

The solution of the H-polarized problem (the magnetic field vector parallel to the edge of the wedge) is simply (2.3) with a plus sign instead of the minus sign:

$$U(\mathbf{r}) = K(\mathbf{r},\mathbf{r}_0) + K(\mathbf{r},\mathbf{r}_0') . \tag{2.3'}$$

It satisfies the P_2 boundary conditions

$$\begin{matrix}\text{boundary}\\ \text{conditions}\end{matrix}\begin{cases} U_+ = U_-, \\ \partial_n U_+ + \partial_n U_- = 0, \end{cases} \text{ on } \Sigma . \tag{2.2'}$$

Thus, it will not be necessary to consider that case explicitly.

The space P_∞ is constructed from an infinite number of copies of P_2. It is the covering space of P_2. The polar angle θ in P_∞ is any real number instead of being modulo 2π. Thus, for any point $\mathbf{r} = (r,\theta)$ in P_2, there are an infinite number of points (preimages) in P_∞ : $(r,\theta + m\Omega)$, $m = 0, \pm 1, \pm 2,\ldots$, whose projections from P_∞ onto P_2 coincide with $\mathbf{r}$.

The desired solution in P_2, $K(\mathbf{r},\mathbf{r}_0)$, is obtained from the corresponding solution in P_∞, $K_\infty(r,\theta;r_0,\theta_0)$, by "folding" it onto P_2; i.e., by summing the fields at all of the preimages:

$$K(\mathbf{r},\mathbf{r}_0) = \sum_{m=-\infty}^{\infty} K_\infty(r,\theta + m\Omega;r_0,\theta_0) . \tag{2.4}$$

One can interpret the image contributions as multiply reflected waves between the boundaries 0 and Ω of P_2. The problem on P_∞ corresponds to one involving a perfectly absorbing wedge. This construction was used in a similar context by Deschamps[17] and by Felsen and Marcuvitz[18] and for quantum mechanical problems by Schulman.[3–5]

III. PATH-INTEGRAL SOLUTION OF A HEAT CONDUCTION EQUATION

The path integral solution of the heat conduction equation

$$\{\partial_\tau - \Delta\}G_F(x,x_0;\tau) = 0 , \tag{3.1}$$

which reduces to $\delta(x - x_0)$ for $\tau = 0$, is reviewed briefly to establish notations. The points x_0 and x are assumed to be in an n-dimensional space $X \equiv \mathbb{R}^n$.

Let γ be a path; i.e., a parametrized arc of a curve in X. It is a map of a segment $[\alpha,\beta]$ of the real axis $\mathbb{R}$ into X and is assumed at least to be continuous. The end points are taken to be $x_0 = \gamma(\alpha)$ and $x = \gamma(\beta)$. Thus,

$$\gamma: [\alpha,\beta] \subset \mathbb{R} \to X: \ \tau \to \gamma(\tau): \ (\alpha,\beta) \mapsto (x_0,x) .$$

Let Γ be the set of continuous paths γ joining x_0 to x in time τ.

The path integral solution of (3.1) is[19]

$$G_F(x,x_0;\tau) = \int_\Gamma F(\gamma)\mathscr{D}\gamma . \tag{3.2}$$

The "value" $F(\gamma)$ assigned to a path $\gamma\in\Gamma$ is taken to be the probability of going from x_0 to x in time τ following the path γ: $F(\gamma) = \exp[-E(\gamma)]$, where the energy of a particle of mass $\frac{1}{2}$ along γ is $E(\gamma) = \frac{1}{4}\int_0^\tau \dot{\gamma}^2\,d\tau$, $\dot{\gamma} = d\gamma/d\tau$ being the velocity of that particle along γ. The quantity $\mathscr{D}\gamma$ is difficult to establish; it represents the measure of the path space. The standard method of giving a meaning to (3.2) is heuristic; the scheme imitates the process that leads to a Riemann integral.

Let each γ be represented by a skeleton γ_s constructed from the $(N+1)$ points $(x_0,x_1,\ldots,x_N = x)$ in the image of γ such that

$$\gamma_s\colon\ (\tau_0 = \alpha,\tau_1,\tau_2,\ldots,\tau_N = \beta)\mapsto(x_0,x_1,x_2,\ldots,x_N = x) .$$

A broken path γ_N can be constructed from these points by associating to each consecutive pair (τ_{j-1},τ_j) mapped into (x_{j-1},x_j) a path segment $\delta_j\gamma$, chosen in a prescribed manner. A standard choice is to make the image of $[\tau_{j-1},\tau_j]$ by $\delta_j\gamma$ into a straight segment described uniformly in τ; i.e.,

$$\delta_j\gamma\colon\ [\tau_{j-1},\tau_j]\to X\colon\ \tau\mapsto x_j = \frac{\delta_j x}{\delta_j\tau}(\tau-\tau_{j-1}) + x_{j-1} .$$

The notation $\delta_j(\cdot)$ designates increments of $(\cdot)$ corresponding to the jth step (or jth segment); e.g., $\delta_j\tau = \tau_j - \tau_{j-1}$. Other choices are possible.

The heuristic definition of the path integral (3.2) is based on approximating each γ by some γ_N; hence, Γ by Γ_N, the set of broken (discrete) paths γ_N. The preceding construction of the set of broken paths Γ_N, which will be referred to as *discretization*, depends on $n(N+1)$ real parameters, provided that the τ_j's are chosen in a systematic manner. For instance, let $\tau_j = \alpha + j\Delta\tau$, where $\Delta\tau = (\beta-\alpha)/N$. One then has the correspondence

$$\gamma_N\leftrightarrow(x_0,x_1,\ldots,x_N)\in\mathbb{R}^{n(N+1)} .$$

Thus, with the fixed end points x_0 and $x_N = x$, the Euclidean measure in $\mathbb{R}^{n(N-1)}$ can be used to define $\mathscr{D}\gamma_N$. In the limit as $N\to\infty$ and $\max\delta_j\tau\to 0$, The definition of the path integral (3.2) becomes

$$\int_\Gamma F(\gamma)\,\mathscr{D}\gamma = \lim_{N\to\infty}\int_{\Gamma_N} F(\gamma_N)\,\mathscr{D}\gamma_N . \tag{3.3}$$

The value $F(\gamma_N)$ assigned to the discretized path γ_N is the product $F(\gamma_N) = \Pi_{j=1}^N F(\delta_j\gamma)$ of functions defined for each of the steps $\delta_j\gamma$ used to construct γ_N. Those functions represent the probability that the particle at x_{j-1} moves to x_j in the time interval from τ_{j-1} to τ_j and are defined as $F(\delta_j\gamma) = \Phi_{2\delta_j\tau}(\delta_j x)$, where

$$\Phi_\sigma(\xi) = (2\pi\sigma)^{-n/2}\exp(-|\xi|^2/2\sigma) .$$

Thus, with the coefficient $A_N = \Pi_{j=1}^N(4\pi\delta_j\tau)^{-n/2}$ and the energy

$$\mathrm{E}(\gamma_N) = \sum_{j=1}^N E(\delta_j\gamma)\equiv\sum_{j=1}^N\frac{(\delta_j x)^2}{4\delta_j\tau} ,$$

the discretization of the path integral (3.2) becomes

$$G_F(x,x_0;\tau) = \lim_{N\to\infty} A_N\int_{\Gamma_N}\exp[-E(\gamma_N)]\,\mathscr{D}\gamma_N\equiv\lim_{N\to\infty}\left(\frac{4\pi\tau}{N}\right)^{-nN/2}\int_{-\infty}^{\infty}\cdots\int\exp\left[-\sum_{j=1}^N E(\delta_j\gamma)\right]dx_1\,dx_2\ldots dx_{N-1} . \tag{3.4}$$

Now consider the polar coordinate form of the path-integral expression (3.4) when $n = 2$. In $\mathbb{R}^2$, the squared distance between the two points $\mathbf{r}_j = (r_j,\theta_j)$ and $\mathbf{r}_{j-1} = (r_{j-1},\theta_{j-1})$ is

$$|\delta_j\mathbf{r}|^2 = r_j^2 + r_{j-1}^2 - 2r_jr_{j-1}\cos(\theta_j-\theta_{j-1})$$

and the measure

$$dx_1\cdots dx_{N-1} = \prod_{j=1}^{N-1} r_j\,dr_j\,d\theta_j .$$

Thus, with $\epsilon = \tau/N$ the expression (3.4) when $n = 2$ can be represented in $\mathbb{R}^2$ as

$$G_F(\mathbf{r},\mathbf{r}_0;\tau) = \lim_{N\to\infty}(4\pi\epsilon)^{-N}\int_{\mathbb{R}^2}\cdots\int\exp\left[-\sum_{j=1}^N\frac{r_j^2+r_{j-1}^2}{4\epsilon}\right]\exp\left[\sum_{j=1}^N\left(\frac{r_jr_{j-1}}{2\epsilon}\right)\cos(\theta_j-\theta_{j-1})\right]\prod_{j=1}^{N-1} r_j\,dr_j\,d\theta_j . \tag{3.5}$$

As shown in Ref. 20, the exact (free-space) solution

$$G_F(\mathbf{r},\mathbf{r}_0;\tau) = (4\pi\tau)^{-1}\exp[-(r^2+r_0^2)/4\tau]\exp[(rr_0/2\tau)\cos(\theta-\theta_0)]\equiv(4\pi\tau)^{-1}\exp[-|\mathbf{r}-\mathbf{r}_0|^2/4\tau] \tag{3.6}$$

is generated from Eq. (3.5). This result will be duplicated from the PIRS point of view in Sect. VII.

IV. PATH-INTEGRAL SOLUTIONS ON P_∞

Returning now to the wedge problem, the propagator $K_\infty(\mathbf{r},\mathbf{r}_0)$, which satisfies on P_∞ the equation

$$\{\Delta + k^2\}K_\infty(\mathbf{r},\mathbf{r}_0) = -\delta(\mathbf{r}-\mathbf{r}_0) , \tag{4.1}$$

is desired. It can be generated by considering the corresponding parabolic equation problem; i.e., the solution $G_\infty(\mathbf{r},\mathbf{r}_0;\tau)$ of the heat conduction equation

$$\{\partial_\tau - \Delta\}G_\infty(\mathbf{r},\mathbf{r}_0;\tau) = 0 , \tag{4.2}$$

which satisfies the initial condition

$$\lim_{\tau\to 0} G_\infty(\mathbf{r},\mathbf{r}_0;\tau) = \delta(\mathbf{r}-\mathbf{r}_0) , \tag{4.3}$$

is related to the solution of (4.1) as

$$K_\infty(\mathbf{r},\mathbf{r}_0)=\int_C d\tau\, e^{k^2\tau}G_\infty(\mathbf{r},\mathbf{r}_0;\tau), \tag{4.4}$$

where C is a contour in the complex plane from $\tau=0$ to infinity. The choice of the contour is independent of k (see Ref. 6).

Following the scheme outlined in the previous section, the path-integral representation of the P_∞ propagator G_∞ is constructed. It is identical to the (free-space) expression (3.5) except that the integrations must now be realized over P_∞-spaces rather than over $\mathbb{R}^2$-spaces. The differences lies in the integration over the angle variables. In the present case, each angle integration must be taken over the infinite inteval $]-\infty,\infty[$ rather than over the finite interval $[0,2\pi]$ used in the free-space example. The resultant P_∞ expression suggests that for its evaluation it would be advantageous to introduce a Fourier transform.

The rotational symmetry of the problem implies that G_∞ will depend only on the angle difference $(\theta-\theta_0)$:

$$G_\infty(\mathbf{r},\mathbf{r}_0;\tau)=G_\infty(r,r_0,\theta-\theta_0;\tau). \tag{4.5}$$

Therefore, the Fourier transform of G_∞ is defined as

$$\begin{aligned}\hat{G}_\lambda(r,r_0;\tau)&=\frac{1}{2\pi}\int_{-\infty}^{\infty}d(\theta-\theta_0)e^{-i\lambda(\theta-\theta_0)}G_\infty(r,r_0,\theta-\theta_0;\tau)\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\theta\, e^{-i\lambda(\theta-\theta_0)}G_\infty(r,r_0;\theta-\theta_0;\tau);\end{aligned} \tag{4.6}$$

its inverse is

$$\begin{aligned}G_\infty(\mathbf{r},\mathbf{r}_0;\tau)&=\int_{-\infty}^{\infty}d\lambda\, e^{i\lambda(\theta-\theta_0)}\hat{G}_\lambda(r,r_0;\tau)\\&\equiv\mathscr{F}(\theta-\theta_0;\lambda)[\hat{G}_\lambda(r,r_0;\tau)].\end{aligned} \tag{4.7}$$

Thus, by considering $\hat{G}_\lambda$ instead of G_∞ directly, one can concentrate on the radial dependence of the propagator. Substituting the pertinent, modified version of Eq. (3.5) into Eq. (4.6) and decomposing the difference $(\theta-\theta_0)$ into

$$\begin{aligned}\theta-\theta_0&=(\theta_N-\theta_{N-1})+(\theta_{N-1}-\theta_{N-2})\\&\quad+\cdots+(\theta_1-\theta_0)\\&=\sum_{j=1}^{N}(\theta_j-\theta_{j-1})\end{aligned}$$

yields

$$\begin{aligned}\hat{G}_\lambda(r,r_0;\tau)&=\lim_{N\to\infty}\frac{1}{2\pi}\int_0^\infty\cdots\int\prod_{j=1}^{N-1}r_j\,dr_j(4\pi\epsilon)^{-N}\\&\quad\times\exp\left[-\sum_{j=1}^{N}\frac{r_j^2+r_{j-1}^2}{4\epsilon}\right]\int_{-\infty}^{\infty}\cdots\int\prod_{j=1}^{N}\Bigg\{d\theta_j\\&\quad\times\exp\left[\frac{r_jr_{j-1}}{2\epsilon}\cos(\theta_j-\theta_{j-1})-i\lambda(\theta_j-\theta_{j-1})\right]\Bigg\}.\end{aligned} \tag{4.8}$$

Note that the additional angle integration with respect to $\theta_N=\theta$ results from the integral in Eq. (4.6). Taking into account the finiteness of the propagator as $r\to 0$, the expression (4.8) gives

$$\hat{G}_\lambda(r,r_0;\tau)=(4\pi\tau)^{-1}\exp[-(r^2+r_0^2)/4\tau]I_{|\lambda|}(rr_0/2\tau). \tag{4.9}$$

The steps leading from Eq. (4.8) to Eq. (4.9) are described in detail in the Appendix.

Inserting the "radial" propagator (4.9) into Eq. (4.7), one obtains the representation

$$\begin{aligned}G_\infty(\mathbf{r},\mathbf{r}_0;\tau)&=(4\pi\tau)^{-1}\exp[-(r^2+r_0^2)/4\tau]\\&\quad\times\int_{-\infty}^{\infty}d\lambda\, e^{i\lambda(\theta-\theta_0)}I_{|\lambda|}\left(\frac{rr_0}{2\tau}\right),\end{aligned} \tag{4.10}$$

for the solution of the heat equation (4.2) on P_∞. Consequently, with Eq. (4.4) the desired solution of Eq. (4.1) on P_∞ is

$$\begin{aligned}K_\infty(\mathbf{r},\mathbf{r}_0)&=\int_{-\infty}^{\infty}d\lambda\, e^{i\lambda(\theta-\theta_0)}\int_C\frac{d\tau}{4\pi\tau}\\&\quad\times\exp\left[k^2\tau-\frac{(r^2+r_0^2)}{4\tau}\right]I_{|\lambda|}\left(\frac{rr_0}{2\tau}\right).\end{aligned} \tag{4.11}$$

Since (see Ref. 21, 8.424.1)

$$\begin{aligned}\frac{1}{\pi i}\int_0^{\eta+i\infty}\exp\left[\frac{1}{2}\left(t-\frac{\xi^2+\zeta^2}{t}\right)\right]I_\nu\left(\frac{\xi\zeta}{t}\right)\frac{dt}{t}\\=J_\nu(\xi)H_\nu^{(1)}(\zeta),\end{aligned}$$

where $\operatorname{Re}\nu>-1$, $\eta>0$, and $|\xi|<|\zeta|$, Eq. (4.11) becomes

$$K_\infty(\mathbf{r},\mathbf{r}_0)=\frac{i}{4}\int_{-\infty}^{\infty}d\lambda\, e^{i\lambda(\theta-\theta_0)}J_{|\lambda|}(kr_<)H_{|\lambda|}^{(1)}(kr_>), \tag{4.12}$$

where $r_>$ is the larger of r and r_0, $r_<$ the smaller. Note that expression (4.12) coincides with the *Riemann surface fundamental solution* defined by Stakgold (see Ref. 22, pp. 270–271).

V. SOLUTION OF THE WEDGE PROBLEM

As noted in Sec. II, the propagator on P_2 from the (real) source point $\mathbf{r}_0$ to the observation point $\mathbf{r}$, $K(\mathbf{r},\mathbf{r}_0)$, associated with the wedge (diffraction) problem, is generated by folding the P_∞-space solution (4.12) onto the P_2-space. The resultant expression has the form

$$\begin{aligned}K(\mathbf{r},\mathbf{r}_0)&=\frac{i}{4}\sum_{m=-\infty}^{\infty}\int_{-\infty}^{\infty}d\lambda\exp[i\lambda(\theta-\theta_0+m\Omega)]\\&\quad\times J_{|\lambda|}(kr_<)H_{|\lambda|}^{(1)}(kr_>).\end{aligned} \tag{5.1}$$

However, using the Poisson summation formula,

$$\begin{aligned}\sum_{m=-\infty}^{\infty}e^{i\lambda m\Omega}&=\sum_{m=-\infty}^{\infty}e^{i2\pi m\beta\lambda}=\sum_{m=-\infty}^{\infty}\delta(\beta\lambda-m)\\&=\frac{1}{\beta}\sum_{m=-\infty}^{\infty}\delta\left(\lambda-\frac{m}{\beta}\right),\end{aligned} \tag{5.2}$$

one obtains

$$\begin{aligned}K(\mathbf{r},\mathbf{r}_0)&=\frac{i}{4\beta}\sum_{m=-\infty}^{\infty}\exp\left[i\frac{m}{\beta}(\theta-\theta_0)\right]\\&\quad\times J_{|m/\beta|}(kr_<)H_{|m/\beta|}^{(1)}(kr_>).\end{aligned} \tag{5.3}$$

This expression can be immediately rewritten as

$$K(\mathbf{r}, \mathbf{r}_0) = \frac{i}{4\beta} \sum_{m=0}^{\infty} \epsilon_m J_{m/\beta}(kr_<) H^{(1)}_{m/\beta}(kr_>) \times \cos\left[\frac{m}{\beta}(\theta - \theta_0)\right], \tag{5.4}$$

where the term

$$\epsilon_m = \begin{cases} 1, & \text{if } m = 0, \\ 2, & \text{if } m \neq 0. \end{cases}$$

Using the results of Ref. 23, it also has the integral representation

$$K(\mathbf{r},\mathbf{r}_0) = \frac{-1}{16\pi} \int_A H_0^{(1)}(kR(\alpha))\, \chi_\beta(\alpha,\theta - \theta_0) d\alpha, \tag{5.5}$$

where the distance $R(\alpha) = [r^2 + r_0^2 - 2rr_0 \cos \alpha]^{1/2}$ and the diffraction coefficient

$$\chi_\beta(\alpha,\psi) = \frac{2}{\beta} \frac{\sin(\alpha/\beta)}{\cos(\alpha/\beta) - \cos(\psi/\beta)}.$$

The path of integration A is shown in Fig. 3.

Consequently, with Eq. (2.3) the total solution of the wedge problem is represented as

$$\begin{aligned} U(\mathbf{r}) &= K(r,\theta;r_0,\theta_0) - K(r,\theta;r_0, -\theta_0) \\ &= \frac{i}{4\beta} \sum_{m=0}^{\infty} \epsilon_m \left\{\cos\left[\frac{m}{\beta}(\theta - \theta_0)\right] - \cos\left[\frac{m}{\beta}(\theta + \theta_0)\right]\right\} \\ &\quad \times J_{m/\beta}(kr_<) H^{(1)}_{m/\beta}(kr_>) \\ &= \frac{i}{2\beta} \sum_{m=0}^{\infty} \epsilon_m \sin\left(\frac{m\theta}{\beta}\right) \sin\left(\frac{m\theta_0}{\beta}\right) \\ &\quad \times J_{m/\beta}(kr_<) H^{(1)}_{m/\beta}(kr_>) \end{aligned} \tag{5.6}$$

or as

$$U(\mathbf{r}) = \frac{-1}{16\pi} \int_A H_0^{(1)}(kR(\alpha)) \times [\chi_\beta(\alpha,\theta - \theta_0) - \chi_\beta(\alpha,\theta + \theta_0)] d\alpha. \tag{5.6$'$}$$

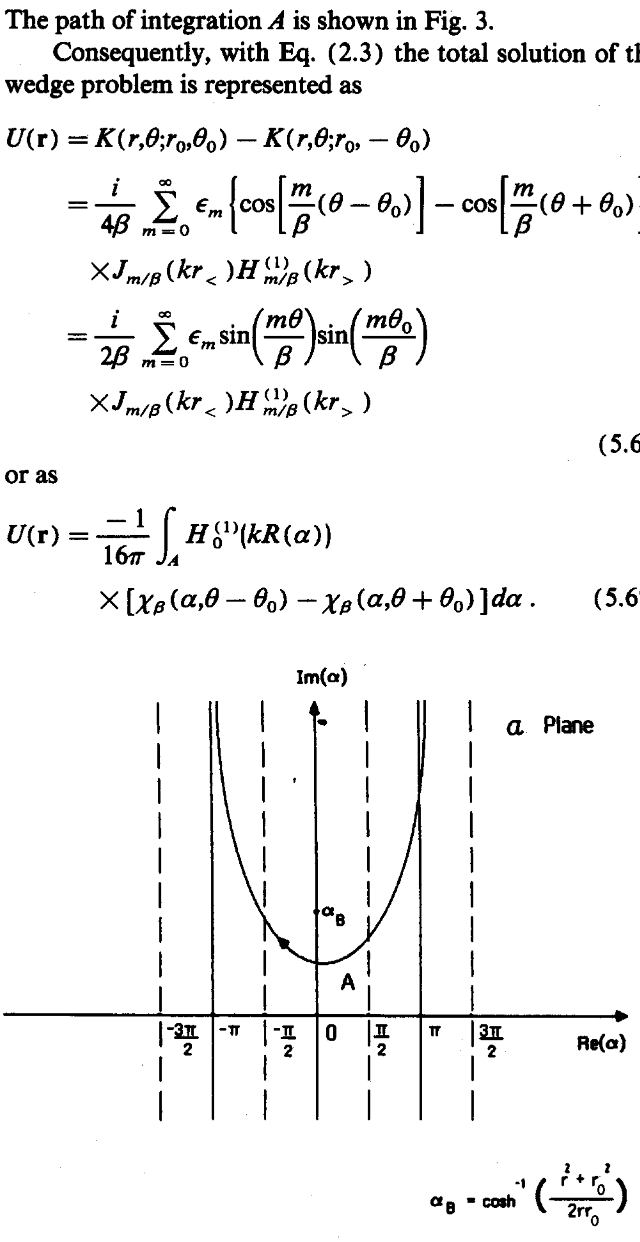


FIG. 3. Path of integration A used for the wedge solution.

These results agree with the known solutions given, for example, in Refs. 18, 21, and 24.

VI. HIGH-FREQUENCY APPROXIMATIONS OF THE WEDGE SOLUTION

The high-frequency (short-wavelength) approximation to the wedge solution (5.6) in the shadow region of the source and its image will be generated from the P_∞ solution (4.12). The GTD results given by Keller[10] are recovered. The analysis is analogous to the one used by Wu[25] to study creeping waves around a circular cylinder.

The P_∞ solution (4.12) can be rewritten as

$$K_\infty(\mathbf{r}, \mathbf{r}_0) = \frac{i}{2} \int_0^\infty d\lambda \cos[\lambda(\theta - \theta_0)] \times J_\lambda(kr_<) H_\lambda^{(1)}(kr_>). \tag{6.1}$$

Asymptotically for $kx \to \infty$ the Bessel and Hankel functions $J_\lambda(kx)$ and $H_\lambda^{(1)}(kx)$ behave as

$$\begin{aligned} \lim_{kx\to\infty} J_\lambda(kx) &\sim (2/\pi kx)^{1/2} \cos[kx - (\lambda + \tfrac{1}{2})\pi/2] \\ &= (2/i)[g(kx)e^{-i\lambda\pi/2} + g^*(kx)e^{i\lambda\pi/2}], \end{aligned}$$

$$\lim_{kx\to\infty} H_\lambda^{(1)}(kx) \sim (4/i) g(kx) e^{-i\lambda\pi/2},$$

where the function

$$g(kx) = (8\pi kx)^{-1/2} \exp[i(kx + \pi/4)]$$

and $g^*(kx)$ is its complex conjugate. Inserting these asymptotic forms into (6.1) and using the relation

$$\int_0^\infty d\lambda\, e^{i\lambda x} = \pi\delta(x) + \frac{i}{x},$$

one obtains the expression

$$\begin{aligned} K_\infty(\mathbf{r}, \mathbf{r}_0) \sim\, & 2g(kr_>)g(kr_<)[1/\psi_+ - 1/\psi_-] \\ & - 2\pi i g(kr_>)g(kr_<)[\delta(\psi_+) + \delta(\psi_-)] \\ & - 4\pi i g(kr_>)g^*(kr_<)\delta[(\psi_+ + \psi_-)/2], \end{aligned} \tag{6.2}$$

where the angles $\psi_\pm = (\theta - \theta_0) \mp \pi = \theta - (\theta_0 \pm \pi)$.

In P_2 the angles $\theta = \theta_0, \theta_0 \pm \pi$ correspond, respectively, to the directions of the source and the shadow boundaries of the source and its image. For a point in the shadow region of the fields incident on the wedge from the source and its image; i.e., for $\theta \epsilon]\theta_0 + \pi, \Omega - (\theta_0 + \pi)[$, these singular directions are not encountered. The asymptotic form of the P_∞ solution then reduces to

$$K_\infty(\mathbf{r},\mathbf{r}_0) \sim 2g(kr_>)g(kr_<)[1/\psi_+ - 1/\psi_-]. \tag{6.3}$$

With (2.4) and (6.3) the propagator $K(\mathbf{r}, \mathbf{r}_0)$ in the shadow region has the asymptotic form

$$K(\mathbf{r}, \mathbf{r}_0) \sim 2g(kr_>)g(kr_<) \times \sum_{m=-\infty}^{\infty} \left[\frac{1}{\psi_+ + m\Omega} - \frac{1}{\psi_- + m\Omega}\right]. \tag{6.4}$$

Since

$$\cot \xi = \frac{1}{\xi} + 2\xi \sum_{m=1}^{\infty} \left[\frac{1}{\xi^2 - (m\pi)^2}\right]$$

and

$$\cot\left(\frac{\phi+\psi}{2}\right)-\cot\left(\frac{\phi-\psi}{2}\right)=-2\frac{\sin\psi}{\cos\psi-\cos\phi},$$

the sum

$$\begin{aligned}2\sum_{m=-\infty}^{\infty}&\left[\frac{1}{\psi_+ + m\Omega}-\frac{1}{\psi_- + m\Omega}\right]\\&=\frac{1}{\beta}\left[\cot\left(\frac{\psi_+}{2\beta}\right)-\cot\left(\frac{\psi_-}{2\beta}\right)\right]\\&=\frac{2}{\beta}\frac{\sin(\pi/\beta)}{\cos(\pi/\beta)-\cos((\theta-\theta_0)/\beta)}\equiv\chi_\beta(\pi,\theta-\theta_0)\end{aligned}\tag{6.5}$$

recovers the wedge diffraction coefficient and (6.4) becomes

$$K(\mathbf{r},\mathbf{r}_0)\sim g(kr_>)\chi_\beta(\pi,\theta-\theta_0)g(kr_<)\,.\tag{6.6}$$

Similarly, the image source contribution is

$$K(\mathbf{r},\mathbf{r}_0')\sim g(kr_>)\chi_\beta(\pi,\theta+\theta_0)g(kr_<)\,.\tag{6.7}$$

Consequently, with (2.3) the asymptotic form of the (E-polarized) wedge solution in the shadow region is

$$U(\mathbf{r})\sim g(kr_>)[\chi_\beta(\pi,\theta-\theta_0)-\chi_\beta(\pi,\theta+\theta_0)]g(kr_<)\,.\tag{6.8}$$

This expression coincides with Keller's GTD result. It represents the effects of the source and image fields interacting with the edge of the wedge. The presence of the additional terms in (6.2) indicates the need in the lit regions to account for the direct, geometrical optics fields in (6.8); i.e., the asymptotic forms of the source and image fields when $\mathbf{r}$ can be reached without interacting with the edge must be included in (6.8).

VII. DISCUSSION

In order to connect the present results with those in the literature, several alternate representations of the PIRS solutions will be considered. They reveal a variety of interesting properties of the PIRS approach.

The point in P_2 that lies on P_- "beneath" the point (r,θ) on P_+ is $(r,\Omega-\theta)$. The value of the P_2 wedge propagator (5.4) at that point recovers the image source contribution

$$K(r,\Omega-\theta;r_0,\theta_0)\equiv K(r,\theta;r_0,-\theta_0)\,.\tag{7.1}$$

Consequently, the wedge solutions (2.3) and (2.3′) can be represented as

$$U(\mathbf{r})=K(r,\theta;r_0,\theta_0)-pK(r,\Omega-\theta;r_0,\theta_0)\,,\tag{7.2}$$

where $p=+1$ for the E-polarized case and $p=-1$ for the H-polarized case. Their satisfaction of the P_2 boundary conditions (2.2) and (2.2′) are easily demonstrated with this expression. Moreover, since $K(r,\theta;r_0,\theta_0)$ and $K(r,\Omega-\theta;r_0\theta_0)$ represent the values of different branches of the P_2 solution at corresponding points, the E- and H-polarized wedge solutions are, respectively, simply a difference and a sum of those values.

Next, consider the sum of the values of the P_∞ solution (4.12) at the points $(r,\theta+m2\pi)$, $m=0,\pm1,\pm2,\ldots$. This sum includes at least one contribution from each P_2 surface in P_∞ and the term $K_\infty(r,\theta+m2\pi;r_0,\theta_0)$ can be viewed as the value of the mth branch of the P_∞ solution at (r,θ). As is readily shown, the free-space propagator $K_F(\mathbf{r},\mathbf{r}_0)$; i.e., the propagator in $\mathbf{R}^2$ between $\mathbf{r}_0$ and $\mathbf{r}$ with no wedge present, is recovered. In particular, with 8.531.2 of Ref. 21 and (5.2) the sum

$$\begin{aligned}\sum_{m=-\infty}^{\infty}&K_\infty(r,\theta+m2\pi;r_0,\theta_0)\\&=\frac{i}{4}\int_{-\infty}^{\infty}d\lambda\exp[i\lambda(\theta-\theta_0)]J_{|\lambda|}(kr_<)\\&\quad\times H^{(1)}_{|\lambda|}(kr_>)\left[\sum_{m=-\infty}^{\infty}e^{i2\pi m\lambda}\right]\\&=\frac{i}{4}\sum_{n=-\infty}^{\infty}e^{in(\theta-\theta_0)}J_{|n|}(kr_<)H^{(1)}_{|n|}(kr_>)\\&=(i/4)H^{(1)}_0(kR(\theta-\theta_0))\equiv K_F(\mathbf{r},\mathbf{r}_0)\,.\end{aligned}\tag{7.3a}$$

This relation illustrates the principle that a symmetric combination of the branches of a multivalued solution to a particular equation such as (2.1) returns a single-valued solution of that equation [See Ref. 7(b), pp. 266–271].

With this result in hand, let us return now to the free-space electromagnetics problem. To account for the enlarged path set, a Riemann space in which each sheet is a replica of $\mathbf{R}^2$; i.e., $P_2\equiv\mathbf{R}^2$, is introduced. The space P_∞ then resembles the spiral staircase surface associated with the logarithm function of complex analysis, and the preimages of (r,θ) are the points $(r,\theta+m2\pi)$, $m=0,\pm1,\pm2,\ldots$. The P_∞ solution $K_\infty(r,\mathbf{r}_0)$ remains (4.12). The folding of K_∞ onto $P^2=\mathbf{R}^2$ given by (7.3a) leads to the exact solution, the free-space propagator $K_F(\mathbf{r},\mathbf{r}_0)$. Similarly, the folding of G_∞ onto $\mathbf{R}^2=P_2$ recovers (3.6), the free-space heat conduction propagator:

$$\begin{aligned}\sum_{m=-\infty}^{\infty}&G_\infty(r,\theta+m2\pi,r_0,\theta_0;\tau)\\&=(4\pi\tau)^{-1}\exp[-(r^2+r_0^2)/4\tau]\\&\quad\times\sum_{m=-\infty}^{\infty}e^{im(\theta-\theta_0)}I_{|m|}\left(\frac{rr_0}{2\tau}\right)\equiv G_F(\mathbf{r},\mathbf{r}_0)\,.\end{aligned}\tag{7.3b}$$

This PIRS description actually provides an alternate representation of the free-space results discussed in Sec. II.

Notice that for the scattering problem where $\beta=+1$, P_+ and P_- are copies of the upper half-plane of $\mathbf{R}^2$ so that P_2 is a double covering of the upper half-plane, not $\mathbf{R}^2$ itself. Thus, even though the preimages of (r,θ) are $(r,\theta+m2\pi)$, $m=0,\pm1,\pm2,\ldots$, and the folding (7.3a) gives $K(\mathbf{r},\mathbf{r}_0)=K_F(\mathbf{r},\mathbf{r}_0)$, an image source is present on P_- and Eq. (5.6) returns the exact solution to the infinite ground plane problem, not the free-space propagator itself.

Comparing the PIRS solutions of the free-space and the wedge problems, the modification of the free-space path set by the presence of the wedge has been modeled simply by constructing the P_∞ space from replicas of the wedge P_2. This modification was responsible for reproducing the diffraction effects. In particular, it led to the evaluation of the P_∞ solution K_∞ at the resultant preimages $(r,\theta+m\Omega)$, $m=0,\pm1,\pm2,\ldots$, of (r,θ), hence, to the propagator (5.4) and the associated image term. This path set modification concept has been used in a companion paper[26] as the basis for a path-integral derivation without discretization of the solution to the diffracting half-plane problem.

Next, the PIRS approach will be connected to several standard quantum mechanical PI methods. This discussion is facilitated by focusing attention on the related results for the heat conduction version of the wedge problem. The P_2-space propagator for the heat conduction problem corresponding to the original diffraction problem is

$$G(\mathbf{r},\mathbf{r}_0;\tau)=\sum_{m=-\infty}^{\infty} G_\infty(r,\theta+m\Omega,r_0,\theta_0;\tau)$$
$$=\sum_{m=-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i\lambda(\theta-\theta_0+m\Omega)}\widehat{G}_\lambda(r,r_0;\tau)\,d\lambda\,. \tag{7.4}$$

Clearly, it is connected to the propagator $K(\mathbf{r},\mathbf{r}_0)$ through the relation

$$K(\mathbf{r},\mathbf{r}_0;\tau)=\int_C e^{k^2t}G(\mathbf{r},\mathbf{r}_0;\tau)d\tau\,.$$

Therefore, with 8.424.1 from Ref. 21, expressions equivalent to (7.4),

$$G(\mathbf{r},\mathbf{r}_0;\tau)$$
$$=\frac{1}{\beta}\sum_{m=-\infty}^{\infty}\exp\Big[i\frac{m}{\beta}(\theta-\theta_0)\Big]\widehat{G}_{m/\beta}(r,r_0;\tau)$$
$$\equiv(4\pi\beta\tau)^{-1}\exp[-(r^2+r_0^2)/4\tau]$$
$$\times\sum_{m=0}^{\infty}\epsilon_m I_{m/\beta}\Big(\frac{rr_0}{2\tau}\Big)\cos\Big[\frac{m}{\beta}(\theta-\theta_0)\Big] \tag{7.5}$$

and

$$G(\mathbf{r},\mathbf{r}_0;\tau)=(i/4\pi)(e^{-(r^2+r_0^2)/4\tau}/4\pi\tau)$$
$$\times\int_A\exp\Big[\frac{rr_0}{2\tau}\cos\alpha\Big]\chi_\beta(\alpha,\theta-\theta_0)\,d\alpha\,, \tag{7.5'}$$

can be extracted from the results presented in Sec. V. They agree with those reported in Ref. 27. These representations of $G(\mathbf{r},\mathbf{r}_0;\tau)$ accommodate several PI interpretations discussed in the literature. Of course, the expressions for the diffraction propagator will acquire similar explanations.

Notice, for instance, that (7.4) can be rewritten as

$$G(\mathbf{r},\mathbf{r}_0;\tau)=\int_{-\infty}^{\infty}\mathscr{G}_\phi(\mathbf{r},\mathbf{r}_0;\tau)d\phi\,, \tag{7.6}$$

where

$$\mathscr{G}_\phi(\mathbf{r},\mathbf{r}_0;\tau)$$
$$=\sum_{m=-\infty}^{\infty}\delta(\phi-[\theta-\theta_0+m\Omega])$$
$$\times\mathscr{F}(\phi;\lambda)[\widehat{G}_\lambda(r,r_0;\tau)]\,, \tag{7.7}$$

and as

$$G(\mathbf{r},\mathbf{r}_0;\tau)=\sum_{m=-\infty}^{\infty}G_m(\mathbf{r},\mathbf{r}_0;\tau)\,, \tag{7.8}$$

where

$$G_m(\mathbf{r},\mathbf{r}_0;\tau)$$
$$=\int_{-\infty}^{\infty}d\lambda\,\exp[i\lambda(\theta-\theta_0+m\Omega)]\widehat{G}_\lambda(r,r_0;\tau)\,. \tag{7.9}$$

The delta function that appears in Eq. (7.7) selects out the probability function (7.9) associated with a particular set of topologically equivalent configurations of the paths with $\phi=(\theta-\theta_0)+m\Omega$. The resultant sum (7.8) extends over all the inequivalent sets contained in the original path set. This explanation has been advocated, for example, by Inomata and Singh.[13] Another interpretation follows Schulman's point of view given, for example, in Ref. 3. The original path set in P_2 can also be decomposed into classes of homotopically equivalent paths labeled by the intersection number, $n(\gamma,\Sigma)$, of their elements γ with Σ. This intersection number is defined as follows. Let Σ_+ and Σ_- be, respectively, the wedge faces $\theta=\beta\pi$ and $\theta=0$ so that $\Sigma=\Sigma_+\cup\Sigma_-$. Let an intersection of a path γ with Σ_+ be positive if γ traverses Σ_+ in the direction from P_+ to P_-, negative if from P_- to P_+, and with Σ_- be positive if the crossing is from P_- to P_+, negative if its from P_+ to P_-. Also let $n_+(\gamma,C)$ and $n_-(\gamma,C)$ be the number of positive and negative crossings of C by γ. Then the intersection number of a path γ connecting $\mathbf{r}_0$ to $\mathbf{r}$ in P_2 is

$$n(\gamma,\Sigma)=[n(\gamma,\Sigma_+)+n(\gamma,\Sigma_-)]/2 \tag{7.10a}$$

where

$$n(\gamma,C)=n_+(\gamma,C)-n_-(\gamma,C)\,. \tag{7.10b}$$

The function G_m then represents the contribution to the propagator from those paths whose intersection number is m. These points of view are equivalent and coincide with the previous preimage description. In particular, the projection onto P_2 of a path connecting $\mathbf{r}_0$ to the preimage $\mathbf{r}_m=(r,\theta+m\Omega)$ of $\mathbf{r}$ coincides with a path γ_m whose intersection number is m. Moreover, since P_2 is isomorphic to $\mathbb{R}^2\setminus\{0\}$, the punctured disk, γ_m is isomorphic to a path in $\mathbb{R}^2\setminus\{0\}$ whose winding number with respect to the origin $\{0\}$ is m. Then mimicking Schulman,[3,4] the term G_m also represents the contribution to the propagator from the paths whose winding number is m.

In Refs. 13 and 14, the path integrals are evaluated directly using the homotopically equivalent path set decomposition. This is accomplished by introducing a constraint into the path integral that distinguishes inequivalent homotopy classes. This "constrained path integral" (CPI) approach realizes a path integral of the form

$$W_\lambda(\mathbf{r},\mathbf{r}_0;\tau)=\int_\Gamma\exp[-S_\lambda(\mathbf{r},\mathbf{r}_0;\tau)]\,\mathscr{D}\gamma\,, \tag{7.11}$$

where the action

$$S_\lambda(\mathbf{r},\mathbf{r}_0;\tau)=\int_0^\tau\Big(\frac{1}{4}\dot{\mathbf{r}}^2+i\lambda\dot\theta\Big)dt\equiv E(\lambda)+i\lambda\int_0^\tau\mathbf{A}\cdot\dot{\mathbf{r}}\,dt\,. \tag{7.12}$$

As noted in Refs. 12–16, the introduction of the linear term $-i\lambda\int_0^\tau\mathbf{A}\cdot\dot{\mathbf{r}}\,dt$, where $\mathbf{A}=(-y,x)/(x^2+y^2)$, in the exponent of (7.11) facilitates the separation of the homotopy classes. The path integral (7.11) is evaluated by discretization; the desired propagator is finally generated through the expression $\int_{-\infty}^{\infty}d\phi\int_{-\infty}^{\infty}d\lambda\,e^{i\lambda\phi}W_\lambda(\mathbf{r},\mathbf{r}_0;\tau)$. Since the linear term in (7.12) is equal to $-i\lambda\int_\gamma d\theta=-i\lambda(\theta-\theta_0)$ and since the constraint and the folding schemes are analogous (as noted above), it is recognized that the PIRS and the

CPI approaches are interrelated. Consequently, it may be possible to extend those problems (quantum mechanical and statistical problems, entangled polymer chains, potential interactions, etc.) to ones involving more general Riemann surfaces like the ones considered here, thus accommodating other physical phenomena. For instance, one obtains an interesting conclusion from the solution of the Aharonov–Bohm problem,[28] which considers quantum mechanical interference effects resulting from potentials in regions where the field is null. Path integral solutions to that problem were considered[4,13–16] from the point of view of electron paths encircling a singular point in a multiply connected space. In particular, the solution to the Aharanov–Bohm problem satisfies on $\mathbb{R}^2\setminus\{0\}$ the Schrödinger equation[28]

$$\{\partial_\tau - (i\hbar/2\mu)[\partial_r^2 + r^{-1}\partial_r + r^{-2}(\partial_\theta - i\alpha)^2]\} \times W(\mathbf{r},\mathbf{r}_0;\tau) = \delta(\mathbf{r}-\mathbf{r}_0)\delta(\tau)\,, \tag{7.13}$$

in the gauge $A_r = 0, A_\theta = \phi/2\pi r$, where ϕ is the "flux" of A through any circuit containing the origin (or equivalently, the flux of the corresponding magnetic field through a surface whose boundary is a circuit) and $\alpha = e\phi/2\pi\hbar c$. It can be represented as

$$W(\mathbf{r},\mathbf{r}_0;\tau) = \sum_{m=-\infty}^{\infty}\int_{-\infty}^{\infty} d\lambda\, \exp[i(\lambda+\alpha)(\theta-\theta_0+2\pi m)] \times \widehat{G}_\lambda(r,r_0;i\hbar\tau/2\mu) = \sum_{m=-\infty}^{\infty} W_n(\mathbf{r},\mathbf{r}_0;\tau)\,, \tag{7.14}$$

which is a variant of the CPI expressions derived in Refs. 13–15. The corresponding P_2-space problem has the solution

$$\widetilde{W}(\mathbf{r},\mathbf{r}_0;\tau) = \sum_{m=-\infty}^{\infty}\int_{-\infty}^{\infty} d\lambda\, \exp[i(\lambda+\alpha)(\theta-\theta_0+m\Omega)] \times \widehat{G}_\lambda(r,r_0;i\hbar\tau/2\mu)\,. \tag{7.15}$$

Interference between the partial propagators W_m and W_n $(m\neq n)$ of (7.14) produces observable interference patterns that depend upon the encircled flux *and* the topological winding number.[15,29,30] On the other hand, quantization of the flux ϕ encircled by the paths can be inferred from the total propagator (7.14) by applying the (two-dimensional) arguments given in Ref. 13. Letting the singular point represent a magnetic monopole with flux $\phi = 4\pi g$ and setting $\mathbf{r} = \mathbf{r}'$, self-consistency requires $2\pi\alpha = \text{integer}\times 2\pi = 2\pi n$ so that the quantization condition derived by Dirac,[31] $g = n(\hbar c/2e)$, is recovered. Similar arguments applied to (7.15) yield $\alpha\Omega = 2\pi n$ or $g = (n/\beta)(\hbar c/2e)$, which means the wedgelike solution corresponds to fractional charge quantization. This result is extended to more general fractions simply by incorporating Riemann surfaces with more sheets. For instance, a Riemann surface P_3 constructed from three copies of P would make $\Omega = 3(\beta\pi)$ and then a choice of $\beta = 2$ would give $g = (n/3)\,(57.5e)$. Thus, the PIRS approach may have some applications in the analysis of quantum field problems involving fractionally charged particles such as quarks.

The special case of the half-plane problem $(\beta = 2)$ leads to another very interesting characteristics of the P_2-space heat conduction and wedge propagators. These half-plane propagators, denoted explicitly by G_2 and K_2, have the forms

$$G_2(\mathbf{r},\mathbf{r}_0;\tau) = (8\pi\tau)^{-1}\exp[-(r^2+r_0^2)/4\tau] \times \sum_{m=0}^{\infty}\epsilon_m I_{m/2}(rr_0/2\tau)\cos\left[m\left(\frac{\theta-\theta_0}{2}\right)\right], \tag{7.16}$$

$$K_2(\mathbf{r},\mathbf{r}_0) = \frac{i}{8}\sum_{m=0}^{\infty}\epsilon_m J_{m/2}(kr_>)H^{(1)}_{m/2}(kr_>) \times\cos[m((\theta-\theta_0)/2)]\,. \tag{7.17}$$

For example, with (4.9), (3.6), and the relations 8.406.1, 8.476.4, and 8.511.4 of Ref. 21,

$$\exp(x\cos\phi) = \sum_{m=0}^{\infty}\epsilon_m I_m(x)\cos(m\phi)\,, \tag{7.18}$$

the expression (7.16) yields the relation

$$\begin{aligned}
G_2(r,\theta+2\pi,r_0,\theta_0;\tau)
&= \frac{1}{2}\sum_{m=0}^{\infty}\epsilon_m\widehat{G}_{m/2}(r,r_0;\tau)\cos\left[m\left(\frac{\theta-\theta_0}{2}\right)+m\pi\right]\\
&= \frac{1}{2}\sum_{\text{even } m}\widehat{G}_{m/2}(r,r_0;\tau)\cos\left[m\left(\frac{\theta-\theta_0}{2}\right)\right]
- \frac{1}{2}\sum_{\text{odd } m}\widehat{G}_{m/2}(r,r_0;\tau)\cos\left[m\left(\frac{\theta-\theta_0}{2}\right)\right]\\
&= \sum_{m=0}^{\infty}\widehat{G}_m(r,r_0;\tau)\cos[m(\theta-\theta_0)] - G_2(r,\theta,r_0,\theta_0;\tau)\\
&= G_F(\mathbf{r},\mathbf{r}_0;\tau) - G_2(r,\theta,r_0,\theta_0;\tau)
\end{aligned}$$

or

$$G_2(r,\theta+2\pi,\mathbf{r}_0;\tau) = -G_2(r,\theta,\mathbf{r}_0;\tau) + G_F(\mathbf{r},\mathbf{r}_0;\tau)\,. \tag{7.19}$$

This also means

$$G_2(r,\theta+4\pi,r_0,\theta_0;\tau) = G_2(r,\theta,r_0,\theta_0;\tau)\,. \tag{7.20}$$

Similarly, the half-plane propagator satisfies

$$K_2(r,\theta+2\pi;\mathbf{r}_0) = -K_2(r,\theta;\mathbf{r}_0) + K_F(\mathbf{r},\mathbf{r}_0)\,, \tag{7.21}$$

$$K_2(r,\theta+4\pi;\mathbf{r}_0) = K_2(r,\theta;\mathbf{r}_0)\,. \tag{7.22}$$

Treating $G_2(r,\theta,2\pi;\mathbf{r}_0)$ and $G_2(r,\theta;\mathbf{r}_0)$ as the values of different branches of G_2 at corresponding points, Eq. (7.19) demonstrates that the half-plane propagator itself exhibits the multivalued solution property:

$$G_2(r,\theta;\mathbf{r}_0) + G_2(r,\theta+2\pi;\mathbf{r}_0) = G_F(\mathbf{r},\mathbf{r}_0)\,. \tag{7.19'}$$

In the half-plane problem Σ is actually a branch line, Σ_+ being its "bottom side" and Σ_- its "top side." Equation (7.20) expresses the continuity of G_2 during the transition through Σ_- from P_- to P_+. On the other hand, evaluating (7.19) at $\theta = \epsilon$, $0<\epsilon\ll 1$, one obtains the transition condition

$$G_2(r,2\pi+\epsilon;\mathbf{r}_0) = -G_2(r,\epsilon;\mathbf{r}_0) + G_F(\mathbf{r},\mathbf{r}_0)\,, \tag{7.23}$$

for the values of G_2 on opposite sides of the branch line Σ, the point (r,ϵ) being in P_+ near Σ_- and $(r,2\pi+\epsilon)$ being in P_-

near Σ_+. Equation (7.23) is also recognized as the transition condition for a Riemann–Hilbert problem[11] for G_2. The -1 coefficient of the transition condition indicates a square root behavior of G_2 near the edge of the half-plane. The corresponding diffraction propagator $K_2(\mathbf{r},\mathbf{r}_0)$ clearly also shares these properties. They are discussed in detail in Ref. 26. Notice, in particular, that requiring K_2 to be bounded at $r=0$ and interpreting the K_2 version of (7.23) as a Riemann–Hilbert problem leads one to Meixner's edge condition[32]

$$\lim_{r\to 0} K_2(\mathbf{r},\mathbf{r}_0) \sim \mathscr{O}(r^{1/2}) .$$

This square root behavior also reinforces the choice of the two-sheeted P_2-space for our analysis,

Similarly, consider the value of G_2 as $\mathbf{r}$ traverses a closed path in the original problem space, where $(r,\theta) = (r,\theta \bmod 2\pi)$, that encloses the edge. If the path has an even winding number, Eq. (7.20) implies that the values of G_2 at the coincident end points of the path are identical, hence that G_2 returns to its original value along a double loop. On the other hand, if its winding number is odd, Eq. (7.19) returns different end-point values. The propagator G_2 does not return to its original value along a single loop but to its negative modified by G_F. Thus, the monodromy group associated with the half-plane problem is $\{1, e^{i\pi} = -1\}$, which is also characteristic of the square root behavior and again indicates the desirability of the two-sheeted P_2-space.

Analogous solution characteristics were utilized by Kadonoff and Kohmoto in their treatment[33] of the two-component spinor correlation function. There, the SMJ (Sato, Miwa, and Jimbo) analysis of the two-dimensional Ising model in terms of the solutions to a two-dimensional version of the Dirac equation and extensions of their analysis were discussed. Since the two-dimensional Dirac and Maxwell equations havae similar forms, the PIRS approach should have applications in statistical mechanics problems as well.

In addition, because potential and heat equation problems are interrelated (probabilistic potential theory[34]), potential problems and the techniques that have been developed to solve them may also prove to be very useful for analyzing the corresponding scattering problems. This concept was first noted by MacDonald.[35] In particular, a P_∞-type analysis of the wedge-potential problem given by Davis and Reitz[36] leads to a solution that is readily connected to the corresponding wedge diffraction solution. Let

$$\Lambda_\pm(\alpha,\psi) = (1/2\pi i)(1/(\psi \pm \alpha)) ,$$

so that

$$\Lambda(\alpha,\psi) = \Lambda_+(\alpha,\psi) - \Lambda_-(\alpha,\psi) ,$$

and let $\mathscr{G}_F[R(\theta-\theta_0)]$ be the free-space Green's function for a particular equation (Helmholtz, heat conduction, and Laplace operators in two or three dimensions). The P_∞-space propagator in any of the corresponding (straight) wedge problems can then be represented in the form

$$K_\infty(\mathbf{r},\mathbf{r}_0) = \int_A \mathscr{G}_F[R(\theta-\theta_0)]\Lambda(\alpha,\theta-\theta_0)d\alpha ; \tag{7.24}$$

hence, the associated P_2-space result is

$$\begin{aligned} K(\mathbf{r},\mathbf{r}_0) &= \sum_{m=-\infty}^{\infty} \int_A \mathscr{G}_A[R(\theta-\theta_0)]\Lambda(\alpha,\theta-\theta_0+m\Omega)d\alpha \\ &= \frac{i}{4\pi}\int_A \mathscr{G}_F[R(\theta-\theta_0)]\chi_\beta(\alpha,\theta-\theta_0)d\alpha . \end{aligned} \tag{7.25}$$

On the other hand, numerical solutions to general potential problems have been constructed based upon path-integral concepts. Generalizations of these schemes to the PIRS point of view would allow solution, for instance, of the curved diffracting wedge problem (see Ref. 37, for example). A coordinate net could be constructed in a P_2-space corresponding to the exterior of the wedge, and the paths and their contributions to the path integral could then be computed numerically in a manner similar to the general potential problem approach. Such a numerical scheme would greatly extend the applicability of the PIRS technique.

Finally, Schulman[38] has remarked that the use of the Riemann surface in connection with path integrals is "an embarrassment to purists." On the contrary, as demonstrated in this paper, the PIRS approach is natural and essential for problems in which boundary surfaces or constraints are present. The RS removes the boundaries or constraints thus allowing the PI to be calculated over a path set having no special restrictions. The RS can then be viewed as containing the PI's original path set information, hence, as arising from a purely path integral context.

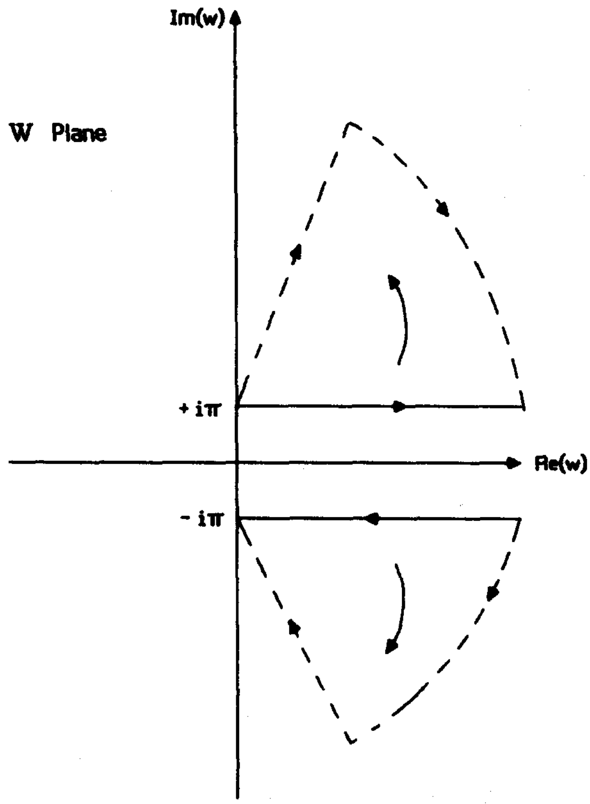


FIG. 4. Deformation of the modified Bessel's function contour.

ACKNOWLEDGMENTS

The author would like to express his deepest appreciation to his former advisor, Professor G. A. Deschamps, for his invaluable comments and suggestions as this paper evolved.

This work was supported in part by National Science Foundation Grant No. NSF-ENG-77-20820 and by the Lawrence Livermore National Laboratory under the auspices of the U. S. Department of Energy under Contract No. W-7405-ENG-48.

APPENDIX: DERIVATION OF THE RADIAL PORTION OF THE P_∞ PROPAGATOR

To generate Eq. (4.9) from Eq. (4.8), first consider the integrations over the angle variables. A typical one has the form

$$f_\lambda(x) = \int_{-\infty}^{\infty} dy \exp(x\cos y - i\lambda y), \tag{A1}$$

where x and λ are real numbers. On the other hand, the modified Bessel's function is given by 6.22.3 in Ref. 39 as

$$I_\lambda(x) = \frac{1}{2\pi i}\int_{\infty - i\pi}^{\infty + i\pi} dw \exp[x\cosh w - \lambda w]. \tag{A2}$$

As shown in Fig. 4, the contour of integration is taken to be $-i\pi + \infty \to -i\pi \to i\pi \to i\pi + \infty$.

Assuming that λ and x are non-negative, the contours parallel to the real axis can be deformed to the ones shown in Fig. 4. The contribution from the arcs at infinity are zero leaving only an integral along the imaginary axis. A change of variables then gives $f_\lambda(x) = 2\pi I_\lambda(x)$. Thus, Eq. (4.8) becomes

$$\hat{G}_\lambda(r,r_0;\tau) = \lim_{N\to\infty}\frac{1}{2\pi}(2\epsilon)^{-N}\int_0^\infty \cdots \int \exp\left[-\sum_{j=1}^{N}\frac{r_j^2 + r_{j-1}^2}{4\epsilon}\right] I_{|\lambda|}\left(\frac{r_1 r_0}{2\epsilon}\right)\cdots I_{|\lambda|}\left(\frac{r_N r_{N-1}}{2\epsilon}\right)\prod_{j=1}^{N-1} r_j\, dr_j. \tag{A3}$$

The non-negativity of the order of the modified Bessel's function has been assured by restricting λ to its absolute value.

Since for $Re(\eta) > -1$ and $Re(\xi) > 0$ (see Ref. 21, 6.633.4),

$$\int_0^\infty \exp(-\xi x^2) I_\eta(ax) I_\eta(bx) x\, dx = \frac{1}{2\xi}\exp\left(\frac{a^2+b^2}{4\xi}\right) I_\eta\left(\frac{ab}{2\xi}\right), \tag{A4}$$

it is readily shown that the jth integration in (A3) yields

$$\begin{aligned}
&\frac{\alpha^{j-1}}{j}\exp\left[\frac{-r_0^2}{j(2\alpha)}\right]\int_0^\infty \exp\left[-\left(\frac{j+1}{j}\right)\frac{r_j^2}{(2\alpha)}\right] I_{|\lambda|}\left(\frac{r_{j+1}r_j}{\alpha}\right) I_{|\lambda|}\left(\frac{r_j r_0}{j\alpha}\right) r_j\, dr_j \\
&= \frac{\alpha^{j-1}}{j}\left(\frac{j\alpha}{j+1}\right)\exp\left[\frac{-r_0^2}{j(2\alpha)} + \frac{r_0^2}{j(j+1)(2\alpha)} + \left(\frac{j}{j+1}\right)\frac{r_{j+1}^2}{(2\alpha)}\right] I_{|\lambda|}\left[\frac{r_0}{j\alpha}\cdot\frac{r_{j+1}}{\alpha}\cdot\frac{j\alpha}{(j+1)}\right] \\
&= \frac{\alpha^j}{j+1}\exp\left[\frac{-r_0^2}{(j+1)(2\alpha)}\right]\exp\left[\left(\frac{j}{j+1}\,\frac{r_{j+1}^2}{(2\alpha)}\right)\right] I_{|\lambda|}\left[\frac{r_{j+1}r_0}{(j+1)\alpha}\right],
\end{aligned} \tag{A5}$$

where $\alpha = 2\epsilon$. Therefore, because $N\alpha = 2N\epsilon \equiv 2\tau$, Eq. (A3) becomes

$$\begin{aligned}
\hat{G}_\lambda(r,r_0;\tau) &= \lim_{N\to\infty}\frac{1}{2\pi}\alpha^{-N}\exp\left(\frac{-r^2}{2\alpha}\right)\left\{\frac{\alpha^{N-1}}{N}\exp\left(\frac{-r_0^2}{2N\alpha}\right)\exp\left[\frac{(N-1)r_N^2}{2N\alpha}\right] I_{|\lambda|}\left(\frac{r_N r_0}{N\alpha}\right)\right\} \\
&= \lim_{N\to\infty}\frac{1}{2\pi N\alpha}\exp\left[\frac{-(r^2+r_0^2)}{2N\alpha}\right] I_{|\lambda|}\left(\frac{rr_0}{N\alpha}\right) = (4\pi\tau)^{-1}\exp\left(\frac{-(r^2+r_0^2)}{4\tau}\right) I_{|\lambda|}\left(\frac{rr_0}{2\tau}\right).
\end{aligned} \tag{A6}$$

A physical explanation of the restriction that λ be non-negative is now apparent. It corresponds precisely to the physical property that the propagator (A6) be finite as $r\to 0$; i.e., $I_{-\lambda}$ is proportional to K_λ, which becomes infinite as its argument nears zero. Note that a generalization of this procedure was developed in Ref. 40 and was used in a similar fashion in 13.

[1] J. S. Dowker, J. Phys. A: Gen. Phys. **5**, 936 (1972).
[2] M. G. G. Laidlaw and C. Morette-DeWitt, Phys. Rev. D **3**, 1375 (1971).
[3] L. Schulman, Phys. Rev. **176**, 1558 (1968).
[4] L. Schulman, J. Math Phys. **12**, 304 (1971).
[5] L. Schulman, *Functional Integration and Its Applications* (Claredon, Oxford, 1975), Chap. 12.
[6] V. S. Buslaev, in *Topics in Mathematical Physics*, edited by Sh. Berman (Consultants Bureau, New York, 1967) (English translation), p. 67.
[7] A. Sommerfeld, (a) Math. Ann. **47**, 317 (1896); (b) *Optics* (Academic, New York, 1954), Sec. 38.
[8] H. S. Carslaw, Proc. London Math. Soc. Ser. 2 **30**, 121 (1899).
[9] S. W. Lee, J. Math. Phys. **19**, 1414 (1978).
[10] J. B. Keller, J. Opt. Soc. Am. **52**, 116 (1962).
[11] F. D. Gakhov, *Boundary Value Problems* (Pergamon, New York, 1966); E. I. Zverovich, Russian Math. Surveys **26**, 117 (1971); R. W. Ziolkowski, SIAM J. Math. Anal. **16**, 358 (1985).
[12] S. F. Edwards, Proc. Phys. Soc. London **91**, 513 (1967).
[13] A. Inomata and V. A. Singh, J. Math. Phys. **19**, 2318 (1978).
[14] C. C. Gerry and V. A. Singh, Phys. Rev. D **20**, 2550 (1979).
[15] C. C. Bernido and A. Inomata, Phys. Lett. A **77**, 394 (1980); J. Math. Phys. **22**, 715 (1981).
[16] F. W. Wiegel, Physica A **109**, 609 (1981).
[17] G. A. Deschamps, "Wedge diffraction," paper presented at the USNC/

URSI-IEEE Meeting, Boulder, Colorado, January, 1975.
[18] L. B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves* (Prentice-Hall, Englewood Cliffs, NJ, 1973).
[19] M. Kac, Trans. Am. Math. Soc. **65**, 1 (1949).
[20] S. F. Edwards and Y. V. Gulyaev, Proc. R. Soc. London Ser. A **279**, 229 (1964).
[21] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1965).
[22] I. Stakgold, *Boundary Value Problems of Mathematical Physics* (Macmillan, New York, 1967), Vol. I.
[23] H. S. Carslaw, Proc. London Math. Soc. Ser. 2 **18**, 291 (1919).
[24] *Electromagnetic and Acoustic Scattering by Simple Shapes*, edited by J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi (North-Holland, Amsterdam, 1969), Chap. 6.
[25] T. T. Wu, Phys. Rev. **104**, 1201 (1956).
[26] R. W. Ziolkowski, UCRL-91607, Lawrence Livermore National Laboratory, 1984.
[27] H. S. Carslaw, Proc. London Math. Soc. Ser. 2 **8**, 365 (1910).
[28] Y. Aharanov and D. Bohm, Phys. Rev. **115**, 485 (1959).
[29] A. Inomata, Phys. Lett. A **95**, 176 (1983).
[30] B. S. Deaver and G. B. Donaldson, Phys. Lett. A **89**, 178 (1982); A. Tonomura, T. Matsuda, R. Suzuki, A. Fukuhara, N. Osakabe, H. Umezaki, J. Endo, K. Shinagawa, Y. Sugita, and H. Fujiwara, Phys. Rev. Lett. **48**, 1443 (1982).
[31] P. A. M. Dirac, Proc. R. Soc. London Ser. A **133**, 60 (1931); Phys. Rev. **74**, 817 (1949).
[32] D. S. Jones, *The Theory of Electromagnetism* (Macmillian, New York, 1964), pp. 566–569.
[33] L. P. Kadanoff and M. Kohmoto, Ann. Phys. **126**, 371 (1980).
[34] R. Hersh and R. J. Griego, Sci. Am. **220** (3), 66 (1969).
[35] H. M. MacDonald, Proc. London Math. Soc. Ser. 2 **14**, 410 (1915).
[36] L. C. Davis and J. R. Reitz, J. Math. Phys. **16**, 1219 (1975).
[37] S. W. Lee and G. A. Deschamps, IEEE Trans. Antennas Propag. **AP-24**, 25 (1976).
[38] L. S. Schulman, *Techniques and Approaches of Path Integration* (Wiley, 1981), pp. 40 and 41.
[39] G. N. Watson, *Theory of Bessel Functions* (Cambridge U. P., London, 1966), 2nd ed.
[40] D. Peak and A. Inomata, J. Math. Phys. **10**, 1422 (1969).

Group-related coherent states

Anton Amann
Laboratorium für Physikalische Chemie, Eidgenössische Technische Hochschule, ETH-Zentrum, CH-8092 Zürich, Switzerland

(Received 12 February 1986; accepted for publication 30 April 1986)

Coherent states defined with respect to an irreducible *ray* representation $u: g \to u_g$, $g \in G$, of an arbitrary locally compact separable group G are examined. It is shown that the following conditions (a)–(d) are equivalent: (a) u admits coherent states, (b) u is square integrable, (c) the W^*-system implemented by u is integrable, and (d) u is a subrepresentation of the left regular c-representation, where c is the respective multiplier of u. Furthermore, the group theoretical background of what is called the "P-representation of observables" associated with coherent states is investigated: It is shown that the P-representation (which corresponds to a covariant semispectral measure) fulfills a certain maximality requirement. The P-representation can be used to represent the quantum system in question on the Hilbert space $L^2(G,dg)$ of square-integrable functions (with respect to Haar measure dg) on the kinematical group G.

I. INTRODUCTION

Depending on the particular situation, different approaches have been used to introduce coherent states in a quantum system: For a harmonic oscillator, e.g., coherent states $|\alpha\rangle$ can equivalently be defined (1) as states that minimize the uncertainty relation, (2) as eigenstates of the destruction operator a, (3) as states that obey the classical motion equation $\langle\alpha|x(t)\alpha\rangle = A \sin(\omega t + \varphi)$, or (4) as states created from the ground state $|0\rangle$ by a unitary operator: $\{\exp[\alpha a^* - \alpha^* a]\}|0\rangle = |\alpha\rangle$. In systems other than the harmonic oscillator it is eventually possible to generalize some of the above notions. Of course, different generalizations need not lead to the same class of states.

If the system in question is characterized by a representation of a kinematical group, a variant of approach (4) above turns out to be particularly useful: Let G denote the respective kinematical group and consider a unitary representation $u: g \to u_g$, $g \in G$, on a Hilbert space $\mathscr{H}$. Then for every fixed vector $\xi_0 \in \mathscr{H}$ coherent states $|g\rangle$ can be defined as $|g\rangle = u_g \xi_0$, $g \in G$. The arbitrary vector ξ_0 is sometimes called a fiducial vector (cf. Ref. 1, p. 21).

Such coherent states were introduced for the first time by Perelomov.[2] There G is a Lie group and u is a proper unitary representation of G. In the present paper a slightly extended frame will be used: The kinematical group G will be an arbitrary *separable* locally compact group and $u: g \to u_g$, $g \in G$, will be an irreducible Borel *ray* representation of G on a *separable* Hilbert space $\mathscr{H}$. This means that the representation property

$$u_{g_1} \circ u_{g_2} = u_{g_1 g_2}, \quad g_1, g_2 \in G,$$

is replaced by

$$u_{g_1} \circ u_{g_2} = c(g_1, g_2) u_{g_1 g_2}, \quad g_1, g_2 \in G,$$

where $c(g_1, g_2)$, $g_1, g_2 \in G$, is a complex number of modulus 1. The mapping

$$c: G \times G \to \mathscr{T} \stackrel{\text{def}}{=} \{z \in \mathbb{C} \mid |z| = 1\}$$

is called the multiplier of u. The representation u is assumed to be Borel in the sense that $G \ni g \to \langle \xi | u_g \xi \rangle \in \mathbb{C}$ is a Borel measurable function for every $\xi \in \mathscr{H}$. Ray representations are necessary in quantum mechanics, since states there are not given by vectors of the Hilbert space, but instead by Hilbert space rays $\mathbb{C} \cdot \eta$, $\eta \in \mathscr{H}$.

The unitaries u_g, $g \in G$, transform the vectors $\xi \in \mathscr{H}$ generating the states. Similarly, one can find symmetries of the algebra of observables $\mathscr{B}(\mathscr{H})$, which consists of all bounded linear operators on the Hilbert space $\mathscr{H}$. Every unitary u_g, $g \in G$, implements a symmetry (an automorphism) of $\mathscr{B}(\mathscr{H})$ defined by $\alpha_g \stackrel{\text{def}}{=} u_g x u_g^*$, $g \in G$, $x \in \mathscr{B}(\mathscr{H})$. Conversely, every automorphism κ of $\mathscr{B}(\mathscr{H})$ is implemented by a unitary v, $\kappa(x) = v x v^*$, $x \in \mathscr{B}(\mathscr{H})$. On the level of the automorphisms α_g the multipliers cancel so that the representation property

$$\alpha_{g_1} \circ \alpha_{g_2} = \alpha_{g_1 g_2}, \quad g_1, g_2 \in G,$$

holds. Here u can be chosen Borel if and only if the mappings $\{G \ni g \to \alpha_g(x)\}$, $x \in \mathscr{B}(\mathscr{H})$, are continuous with respect to the σ-weak (equivalently the weak, strong, σ-strong) topology on $\mathscr{B}(\mathscr{H})$. Thus u is irreducible, i.e., describes an elementary quantum system, if and only if α is ergodic, that is, if

$$\alpha_g(x) = x, \quad \forall g \in G,$$

implies $x = c \cdot 1$, $c \in \mathbb{C}$ (see Ref. 3, 67.2).

This algebraic approach can be extended considerably: Instead of $\mathscr{B}(\mathscr{H})$ one can take an arbitrary W^*-algebra $\mathscr{M}$. $\mathscr{M}$ can again support a representation α of a kinematical group G. The triple $(\mathscr{M}, G, \alpha)$ is then called a W^*-system. The W^*-formalism is broad enough to comprehend both quantum systems and classical systems. In the latter case $\mathscr{M}$ is given as the commutative algebra $L_\infty(\Omega)$ of functions on the phase space Ω of classical mechanics. Similarly as above a (transitive, i.e., elementary) group representation on the phase space Ω implements an (ergodic) representation α on $L_\infty(\Omega)$. Note that the W^*-formalism can also describe infinite systems and systems with both classical and quantum properties. A comprehensive review of the algebraic formalism from the physical point of view is given by Primas,[4] a short introduction into its group theoretical aspects can be found in Chap. 1 of Ref. 5.

C- and W*-algebras:* A C^*-*algebra* is a *-algebra that is *-isomorphic to a norm-closed algebra of operators on a Hilbert space. A W^*-*algebra* is a *-algebra that is *-isomorphic to a *-algebra $\mathscr{R} \subseteq \mathscr{B}(\mathscr{H})$ of operators on a Hilbert space $\mathscr{H}$ fulfilling $\mathscr{R} = (\mathscr{R}')'$, where the commutant $\mathscr{S}'$ of a set of operators $\mathscr{S}$ on $\mathscr{H}$ is defined as

$$\mathscr{S}' \overset{\text{def}}{=} \{x \in \mathscr{B}(\mathscr{H}) | xy = yx\,, \quad \forall y \in \mathscr{S}\}\,.$$

Every W^*-algebra is a C^*-algebra but not conversely. W^*-algebras can be characterized intrinsically: A C^*-algebra $\mathscr{M}$ is a W^*-algebra iff it is the dual of a Banach space $\mathscr{M}_*$, where $\mathscr{M}_*$ is called the predual of $\mathscr{M}$.

Every W^*-algebra $\mathscr{M}$ contains a unit element 1. A *state* ω on $\mathscr{M}$ is a positive linear and normalized $[\omega(1) = 1]$ mapping ω: $\mathscr{M} \to \mathbb{C}$. A state ω on $\mathscr{M}$ is called *normal* (or σ-weakly continuous) if there is an element $a \in \mathscr{M}_*$ such that $\omega(x) = x(a)$ holds for all operators $x \in \mathscr{M}$. Equivalently, ω is normal iff $\sup_\beta \omega(x_\beta) = \omega(\sup_\beta x_\beta)$ holds for every bounded increasing net $(x_\beta)_{\beta \in I}$ of positive operators from $\mathscr{M}$.

If a physical system is described by the "algebra of observables" $\mathscr{M}$, then its respective classical properties correspond to nontrivial elements of the *center*

$$\mathscr{Z}(\mathscr{M}) \overset{\text{def}}{=} \{x \in \mathscr{M} | xy = yx\,, \ \forall y \in \mathscr{M}\}\,.$$

If no classical properties exist, i.e., if $\mathscr{Z}(\mathscr{M}) = \mathbb{C} \cdot 1$, then $\mathscr{M}$ is called a *factor*.

The definition of coherent states as given by Perelomov[2] incorporates another important ingredient. It is required that the integral $\int_G |g\rangle\langle g| dg$ exists (cf. Ref. 2 and Ref. 1, p. 5). Here dg denotes the left-invariant Haar measure on the group G. Existence of $\int_G |g\rangle\langle g| dg$ means that there is an operator $x \in \mathscr{B}(\mathscr{H})$, such that

$$\langle \xi | x \xi \rangle = \int_G \langle \xi | g \rangle \langle g | \xi \rangle dg$$

holds for all $\xi \in \mathscr{H}$. Irreducibility of u then implies that $\int_G |g\rangle\langle g| dg = x$ is a multiple of the identity operator 1. Appropriate normalization (of the Haar measure or the fiducial vector ξ_0) then results in

$$\int_G |g\rangle\langle g| dg = 1\,. \tag{1}$$

The resolution of unity (1) is sometimes referred to as the completeness property of the coherent states $|g\rangle$, $g \in G$.

If G is not compact, i.e., the Haar measure dg is not finite, condition (1) is a severe restriction. On the Hilbert space level it implies that there exist nonvanishing vectors ξ and η such that

$$\int_G |\langle u_g \eta | \xi \rangle|^2 \, dg \tag{2}$$

exists. An irreducible ray representation with this property is called *square integrable*. In the context of coherent states square-integrable proper unitary representations have been studied, e.g., in Refs. 6 and 7. Note that condition (1) is not so much a restriction for the fiducial state (this may be the case, too), but a restriction for the unitary ray representation u.

On the algebraic level (1) is equivalent to the existence of an atomic projection $p \in \mathscr{B}(\mathscr{H})$, namely $p = |\xi_0\rangle\langle\xi_0|$, such that

$$\int_G \alpha_g(p) dg < \infty\,. \tag{3}$$

Incidentally, existence of the integral (3) in the weak sense is equivalent to the existence of the supremum $\sup_K \int_K \alpha_g(p) dg$, where K runs over all compact subsets of G. Since α is ergodic, $\int_G \alpha_g(p) dg$ is a multiple of the identity operator 1. If (3) holds for some atomic projection p, the ergodic W^*-system $(\mathscr{B}(\mathscr{H}), G, \{\alpha_g(\cdot) = u_g \cdot u_g^* | g \in G\})$ is called *integrable*. It is the ergodicity of α, which then implies (see Ref. 8, Lemma 1.4) that

$$n_\alpha \overset{\text{def}}{=} \left\{ y \in \mathscr{B}(\mathscr{H}) \,\middle|\, \int_G \alpha_g(y^* y) dg < \infty \right\}$$

is σ-weakly dense in $\mathscr{B}(\mathscr{H})$, which is the usual definition of integrability of a W^*-system as given in Ref. 9. Conversely, if n_α is dense in $\mathscr{B}(\mathscr{H})$, it contains an atomic projection p (cf. the proof of Theorem 2) such that

$$\int_G \alpha_g(p) dg = \int_G \alpha_g(p^* p) dg < \infty$$

holds true.

Thus coherent states do exist with respect to an irreducible ray representation u if and only if the associated W^*-system $(\mathscr{B}(\mathscr{H}), G, \{\alpha_g(\cdot) = u_g \cdot u_g^* | g \in G\})$ is integrable. Assume now this to be the case and set $p = |\xi_0\rangle\langle\xi_0|$. Then for every bounded (Borel-measurable) function m: $G \to \mathbb{C}$ on the group G there exists an operator $\varphi_p(m) \in \mathscr{B}(\mathscr{H})$

$$\varphi_p(m) \overset{\text{def}}{=} \int_G \alpha_g(p) m(g) dg = \int_G |g\rangle\langle g| m(g) dg\,. \tag{4}$$

The operator $\varphi_p(m)$ is then said to admit a diagonal representation or P-representation (cf. Ref. 1, p. 13). Such a representation has been used, e.g., for certain Hamiltonians (cf. Ref. 1, p. 69), but may incorporate much more general observables,[5] for example, position and momentum operators (if G is the group $\mathbb{R}^{2n}$, $n \in \mathbb{N}$) or spin operators [if G is the rotation group SO(3)]. A large part of the theory of integrable systems can then be done on the level of functions on the kinematical group. It is worth remarking that m: $G \to \mathbb{C}$ need not be a bounded function and correspondingly $\varphi_p(m)$ need not be a bounded operator (cf. Ref. 10, Chap. II).

φ_p maps functions on G into operators on the Hilbert space $\mathscr{H}$. This paper aims at giving a precise group-theoretical sense to φ_p (Sec. II) and at studying its properties (Sec. III). Section IV will be devoted to the question under which circumstances an irreducible ray representation implements an integrable W^*-system, i.e., admits coherent states. In the final section it will be shown that the quantum and classical theories of a certain kinematical group G can be represented on one and the same Hilbert space $L^2(G, dg)$.

II. COVARIANT EMBEDDINGS OF $L_\infty(G)$ INTO W^*-SYSTEMS

In the following $L_\infty(G) \overset{\text{def}}{=} L_\infty(G, dg)$ will denote the W^*-algebra of all essentially bounded Borel measurable complex-valued functions on a separable locally compact group G. The action $\mathrm{Ad}\,\lambda$ of G on $L_\infty(G)$ is defined as

$$(\mathrm{Ad}\,\lambda(g)m)(s) \stackrel{\mathrm{def}}{=} m(g^{-1}s)\,,\quad g,s\in G\,.$$

More precisely, the elements of $L_\infty(G)$ are classes of functions differing only on a dg-null set. The norm $\|[m]\|$ of such a class of functions $[m]$ is defined as the infimum

$$\|[m]\| \stackrel{\mathrm{def}}{=} \inf\{\|n\|_{\sup}\,|\,n\in[m]\}\,,$$

where

$$\|n\|_{\sup} = \operatorname*{supremum}_{g\in G}|n(g)|\,.$$

$L_\infty(G)$ is a W^*-algebra with the predual $L_1(G,dg)$ of all integrable Borel measurable function on G.

Covariant embeddings of $L_\infty(G)$ into a W^*-system[5]*:* Let $(\mathscr{M},G,\alpha)$ be a W^*-system. Then a (positive, normalized) *covariant embedding* of $L_\infty(G)$ into $(\mathscr{M},G,\alpha)$ is a mapping $\chi: L_\infty(G)\to\mathscr{M}$ with the properties

(i) χ is linear: $\chi(m+\lambda n)=\chi(m)+\lambda\chi(n)\,,\quad m,n\in L_\infty(G)\,,\quad \lambda\in\mathbb{C}\,,$

(ii) χ is positive: $\chi(m)\geqslant 0$ if $m\in L_\infty(G)_+ \stackrel{\mathrm{def}}{=} \{m\in L_\infty(G)\,|\,m\geqslant 0\}\,,$

(iii) χ is normalized: $\chi(1)=1\,,$

(iv) χ is covariant: $\alpha_g(\chi(m))=\chi(\mathrm{Ad}\,\lambda(g)m)\,,\quad m\in L_\infty(G)\,,\quad g\in G\,.$

A covariant embedding $\chi: L_\infty(G)\to\mathscr{M}$ is called *normal* (or σ-weakly continuous) if for every normal state Ψ on $\mathscr{M}$, the state $\Psi\circ\chi$ on $L_\infty(G)$ is normal.

A covariant normal embedding $\chi: L_\infty(G)\to\mathscr{M}$ is essentially the same as a covariant semispectral measure.[6] In the latter case not all elements from $L_\infty(G)$ are considered but just characteristic functions of Borel subsets of G. Covariant semispectral measures cannot just be used for introducing and studying observables,[5] but also for the discussion of their measurement.[11,12]

Observation 1: Consider an ergodic W^*-system $(\mathscr{M},G,\alpha)$ and a positive operator $x\in\mathscr{M}$ such that $\int_G \alpha_g(x)dg=1$. Then

$$\varphi_x:\ L_\infty(G)\to\mathscr{M}\,,$$

$$\varphi_x:\ m\to\int_G \alpha_g(x)m(g)dg\,,\quad m\in L_\infty(G)\,,$$

is a normal covariant embedding of $L_\infty(G)$ into $(\mathscr{M},G,\alpha)$ (Ref. 5, Lemma II.1).

Observation 2: An ergodic W^*-system $(\mathscr{M},G,\alpha)$ admits a normal covariant embedding $\chi: L_\infty(G)\to\mathscr{M}$ if and only if $(\mathscr{M},G,\alpha)$ is integrable (Ref. 5, Theorem II.2).

The particular normal covariant embedding

$$\varphi_p:\ m\to\int_G |g\rangle\langle g|m(g)dg=\int_G \alpha_g(p)m(g)dg\,,$$
$$m\in L_\infty(G)$$
$$(p=|e\rangle\langle e|\,,\quad e \text{ is the neutral element of } G)$$

has been introduced at the end of Sec. I. Of course, it can happen that coherent states $|g_1\rangle$ and $|g_2\rangle$, $g_1\neq g_2$, are identical or differ just by a complex number of modulus 1. Accordingly, the closed subgroup

$$H\stackrel{\mathrm{def}}{=}\{h\in G\,|\,\alpha_h(p)=p\}=\{h\in G\,|\,\mathbb{C}\cdot|h\rangle=\mathbb{C}\cdot|e\rangle\}$$

may be nontrivial, i.e., $H\neq\{e\}$. The integrability condition (1) then implies that H is a *compact* subgroup of G and furthermore

$$\varphi_p(\mathrm{Ad}\,\rho(h)m)=\varphi_p(m)\,,\quad m\in L_\infty(G)\,,\quad h\in H\,,$$

where $(\mathrm{Ad}\,\rho(h)m)(s)\stackrel{\mathrm{def}}{=}m(sh)$, $s,h\in G$. If dh denotes a normalized Haar measure on the compact group H, the relation

$$\varphi_p\left\{\int_H(\mathrm{Ad}\,\rho(h)m)dh\right\}=\int_H\varphi_p(\mathrm{Ad}\,\rho(h)m)dh$$
$$=\varphi_p(m)\,,\quad m\in L_\infty(G)\,,$$

holds. $\int_H(\mathrm{Ad}\,\rho(h)m)dh$ can be considered as a function on the left coset space G/H and consequently $\varphi_p: L_\infty(G)\to\mathscr{B}(\mathscr{H})$ can then be replaced by a covariant mapping (cf. Refs. 2 and 6)

$$\tilde\varphi_p:\ L_\infty(G/H)\to\mathscr{B}(\mathscr{H})\,,$$

$$\tilde\varphi_p(\tilde m)\stackrel{\mathrm{def}}{=}\varphi_p(\iota(\tilde m)),\quad \tilde m\in L_\infty(G/H)\,,$$

$$\iota(\tilde m)(s)\stackrel{\mathrm{def}}{=}\tilde m(sH)\,.$$

Here $L_\infty(G/H)$ is defined with respect to the unique G-invariant measure class on G/H.

A normal covariant embedding $\chi: L_\infty(G)\to\mathscr{M}$ may be trivial, i.e., $\chi(L_\infty(G))=\mathbb{C}\cdot 1$. However, to be of interest, $\chi(L_\infty(G))$ should exhaust the elements of $\mathscr{M}$ as much as possible. To give a precise definition of this, consider two normal covariant embeddings χ_1 and χ_2 of $L_\infty(G)$ into a W^*-system $(\mathscr{M},G,\alpha)$. Then χ_1 is said to be contained in χ_2, $\chi_1\subseteq\chi_2$, if $\chi_1(L_\infty(G)_+)\subseteq\chi_2(L_\infty(G)_+)$ holds. A normal covariant embedding $\chi: L_\infty(G)\to\mathscr{M}$ is called *maximal* if $\chi\subseteq\varphi$ implies $\chi(L_\infty(G)_+)=\varphi(L_\infty(G)_+)$, for every other normal covariant embedding φ. Maximality is opposed to triviality and the question arises which position is held in this hierarchy by the normal covariant embeddings associated to coherent states (see Sec. III). Existence of maximal embeddings in arbitrary ergodic W^*-systems is settled by the following observation.

Observation 3: (See Ref. 5, Theorem II.7.) Let $(\mathcal{M},G,\alpha)$ be an ergodic W^*-system. Then the following conditions are equivalent: (i) $(\mathcal{M},G,\alpha)$ is integrable, (ii) there exists a normal covariant embedding of $L_\infty(G)$ into $(\mathcal{M},G,\alpha)$, and (iii) there exists a *maximal* normal covariant embedding of $L_\infty(G)$ into $(\mathcal{M},G,\alpha)$.

Observation 4: (See Ref. 5, Theorem II.6.) Let χ_1 and χ_2 be normal covariant embeddings of $L_\infty(G)$ into an ergodic W^*-system $(\mathcal{M},G,\alpha)$ such that $\chi_1(L_\infty(G)_+) = \chi_2(L_\infty(G)_+)$. Then there exists an element $g_0 \in G$, such that $\chi_1(m) = \chi_2(\mathrm{Ad}\,\rho(g_0)m)$, $\forall m \in L_\infty(G)$.

III. COVARIANT EMBEDDINGS ASSOCIATED WITH COHERENT STATES

The main result of this chapter is Theorem 4, stating that normal covariant embeddings of the form $m \to \int_G m(s)|s\rangle\langle s|ds$, $m \in L_\infty(G)$, are maximal with respect to the ordering $\subseteq$ introduced in Sec. II. For the proof of this fact the following lemmas are needed.

Lemma 1: A normal covariant embedding χ of $L_\infty(G)$ into an ergodic W^*-system $(\mathcal{M},G,\alpha)$ is of the form φ_x for a suitable positive operator $x \in \mathcal{M}$ if and only if there exists a constant $N > 0$ such that

$$\|\chi(f)\| < \left\{ \int_G f(g)dg \right\} \cdot N$$

holds for all positive elements f from $\{L^1(G) \cap L_\infty(G)\}$. For φ_x this constant N can be chosen to be $\|x\|$.

The proof of this lemma is given in the Appendix.

Lemma 2: Let $u: g \to u_g$, $g \in G$, be an irreducible ray representation on a Hilbert space $\mathcal{H}$ and consider the representation

$$\alpha_g(\cdot) \overset{\mathrm{def}}{=} u_g \cdot u_g^*, \quad g \in G,$$

on $\mathcal{B}(\mathcal{H})$. Let furthermore $x \in \mathcal{B}(\mathcal{H})_+$ be a positive operator such that $\int_G \alpha_g(x)dg = 1$. Then the associated normal covariant embedding

$$\varphi_x: f \to \int_G \alpha_g(x) f(g)dg, \quad f \in L_\infty(G),$$

is extremal within the set of all normal covariant embeddings of $L_\infty(G)$ into $(\mathcal{B}(\mathcal{H}),G,\alpha)$ if and only if x is a multiple of an atomic projection.

Proof: If x is not a multiple of an atomic projection, φ_x is clearly not extremal even within covariant embeddings of the form φ_y, $y \in \mathcal{B}(\mathcal{H})_+$, $\int_G \alpha_g(y)dg = 1$. Conversely, let x be the multiple of an atomic projection, $\int_G \alpha_g(x)dg = 1$, and suppose $\varphi_x = \lambda\chi_1 + (1-\lambda)\chi_2$, $\lambda \in (0,1)$, where χ_i, $i = 1,2$, are normal covariant embeddings:

$$0 < \lambda\chi_1(f) < \varphi_x(f), \quad \forall f \in \{L^1(G) \cap L_\infty(G)\}_+$$
$$\Rightarrow \|\chi_1(f)\| < \frac{1}{\lambda}\|\varphi_x(f)\| < \frac{\|x\|}{\lambda}\left\{ \int_G f(g)dg \right\}.$$

From Lemma 1 it is now inferred that χ_1 (and similarly χ_2) is of the form $\chi_1 = \varphi_{x_1}$ ($\chi_2 = \varphi_{x_2}$), where x_1 and x_2 are positive operators from $\mathcal{B}(\mathcal{H})$ such that $\int_G \alpha_g(x_i)dg = 1$, $i = 1,2$:

$$\Rightarrow \varphi_x = \lambda\varphi_{x_1} + (1-\lambda)\varphi_{x_2},$$
$$\Rightarrow \int_G \Psi(\alpha_g(x - \lambda x_1 - (1-\lambda)x_2)) f(g) = 0, \quad \forall \Psi \in \mathcal{M}_*, \quad \forall f \in L_\infty(G),$$
$$\Rightarrow x = \lambda x_1 + (1-\lambda)x_2.$$

Since x is a multiple of an atomic projection p, $x = \mu p$, one infers that $x_i = p x_i p = \mu_i p$, $\mu_i \in \mathbb{R}_+$, $i = 1,2$. Then

$$\int_G \alpha_g(x)dg = \int_G \alpha_g(x_i)dg, \quad i = 1,2,$$

implies $x = x_1 = x_2$ and $\chi = \chi_1 = \chi_2$. Q.E.D.

Theorem 1: Let $u: g \to u_g$, $g \in G$, be an irreducible ray representation on a Hilbert space $\mathcal{H}$ and consider the representation $\alpha_g(\cdot) = u_g \cdot u_g^*$, $g \in G$, on $\mathcal{B}(\mathcal{H})$. Let $x \in \mathcal{B}(\mathcal{H})_+$ be a multiple of an atomic projection such that $\int_G \alpha_g(x)dg = 1$. Then the associated covariant embedding $\varphi_x: L_\infty(G) \to \mathcal{B}(\mathcal{H})$ is *maximal*, i.e., every normal covariant embedding χ with the property $\varphi_x(L_\infty(G)_+) \subseteq \chi(L_\infty(G)_+)$ fulfills $\varphi_x(L_\infty(G)_+) = \chi(L_\infty(G)_+)$. Furthermore, the latter identity implies that there exists an element $g_0 \in G$, such that $\chi = \varphi_{\alpha_{g_0}(\Delta(g_0)x)}$.

Proof: The normal covariant embedding φ_x is extremal within the set of all covariant embeddings of $L_\infty(G)$ into $(\mathcal{B}(\mathcal{H}),G,\alpha)$ (use Lemma 2 and Ref. 5, proof of Theorem II.5). The maximality then follows from Ref. 5, Theorem II.6. Here $\varphi_x(L_\infty(G)_+) = \chi(L_\infty(G)_+)$ implies (see Ref. 5, Theorem II.6) that there exists an element $g_0 \in G$, such that $\varphi_x(\mathrm{Ad}\,\rho(g_0^{-1})m) = \chi(m)$, $\forall m \in L_\infty(G)$. The last assertion of the theorem then follows from

$$\begin{aligned}
\varphi_x(\mathrm{Ad}\,\rho(g_0^{-1})m) &= \int_G \alpha_g(x) m(g g_0^{-1})dg \\
&\overset{g' = g g_0^{-1}}{=} \int_G \alpha_{g' g_0}(x)\Delta(g_0) m(g')dg' \\
&= \int_G \alpha_g(\Delta(g_0)\alpha_{g_0}(x)) m(g)dg \\
&= \varphi_{\alpha_{g_0}(\Delta(g_0)x)}(m), \quad \forall m \in L_\infty(G).
\end{aligned}$$

Q.E.D.

Remark: If x and y are multiples of atomic projections such that no $g \in G$ exists with $x = \alpha_g(\Delta(g)y)$, the associated maximal normal covariant embeddings φ_x and φ_y have different range, $\varphi_x(L_\infty(G)_+) \neq \varphi_y(L_\infty(G)_+)$. A simple example, based on a commutative group of four elements, is given in Chap. III of Ref. 10. The maximal normal covariant embeddings in this system can be studied in a geometric manner. Additional criteria can be found there to mark out certain maximal normal covariant embeddings and thus certain atomic projections or states of the system.

On the other hand there exist systems $(\mathcal{B}(\mathcal{H}),G,\{\alpha_g(\cdot) = u_g \cdot u_g^*, g \in G\})$ that essentially admit only one maximal normal covariant embedding. A typical example is given by the spin-$\frac{1}{2}$ irreducible ray representation of the rotation group SO(3) (see Ref. 10, Chap. III).

Conjecture: If $u: g \to u_g$, $g \in G$, is an irreducible ray representation on a Hilbert space $\mathcal{H}$, every maximal normal covariant embedding of $L_\infty(G)$ into the W^*-system

$(\mathscr{B}(\mathscr{H}),G,\{\alpha_g(\cdot)=u_g\cdot u_g^*,\ g\in G\})$ is of the form φ_x: $m\to\int_G \alpha_g(x)m(g)dg$, $m\in L_\infty(G)$, where x is a multiple of an atomic projection.

IV. EXISTENCE OF COHERENT STATES

The starting point of this chapter is the observation that square integrability of an irreducible unitary ray representation implies the integrability of the associated W^*-system and vice versa. Note that the concept of integrable unitary representations as presented in Ref. 13, 14.5 is not useful in the present context, since ray representations u: $g\to u_g$, $g\in G$, and $\tilde{u}$: $g\to d(g)u_g$, $|d(g)|=1$, $g\in G$, implement the same actions on the algebra $\mathscr{B}(\mathscr{H})$ and are considered as being equivalent. The following definition will be needed.

Definition: Let c: $G\times G\to\mathscr{T}$ be a multiplier. Then the ray representation λ^c on the Hilbert space $L^2(G,dg)$ defined by

$$(\lambda^c(g)\xi(x)) \overset{\text{def}}{=} c(g,g^{-1}s)\xi(g^{-1}s)\,,$$

$$\xi\in L^2(G,dg)\,,\quad s,g\in G\,,$$

is called the left regular c-representation of G.

Theorem 2: Consider an irreducible ray representation u: $g\to u_g$, $g\in G$, on a Hilbert space $\mathscr{H}$. Then u is square integrable if and only if the associated W^*-system $(\mathscr{B}(\mathscr{H}),G,\{\alpha_g(x)\overset{\text{def}}{=}u_gxu_g^*,\ x\in\mathscr{B}(\mathscr{H})|g\in G\})$ is integrable.

Proof: $\Rightarrow$: In (a) and (b) part of the proof of Ref. 14, Theorems 2 and 3 is mimicked.

(a) Suppose u is square integrable, i.e., there exists vectors ξ_0 and η of norm 1 such that $\int_G|\langle u_s\eta|\xi_0\rangle|^2\,ds<\infty$. Then the operator V_η, defined by

$$V_\eta\xi \overset{\text{def}}{=} \{G\ni s\to\langle u_s\eta|\xi\rangle\}\in L^2(G)\,,\quad \xi\in D(V_\eta)\,,$$

$$D(V_\eta) \overset{\text{def}}{=} \left\{\xi\in\mathscr{H}\,\middle|\,\int_G|\langle u_s\eta|\xi\rangle|^2\,ds<\infty\right\},$$

has a nontrivial domain of definition $D(V_\eta)$ and is closed.

(b) $\lambda^c(g)V_\eta\xi=V_\eta u_g\xi$, $\forall g\in G$, $\forall\xi\in D(V_\eta)$:

$$\begin{aligned}
&(\lambda^c(g)V_\eta\xi)(s)\\
&\quad=c(g,g^{-1}s)(V_\eta\xi)(g^{-1}s)\\
&\quad=c(g,g^{-1}s)\langle u_{g^{-1}s}\,\eta|\xi\rangle\\
&\quad=c(g,g^{-1}s)\langle u_gu_g^*u_{g^{-1}s}\,\eta|\xi\rangle\\
&\quad=c(g,g^{-1}s)\langle c(g,g^{-1}s)u_s\eta|u_g\xi\rangle=\langle u_s\eta|u_g\xi\rangle\\
&\quad=(V_\eta u_g\xi)(s)\,,\quad s,g\in G\,.
\end{aligned}\tag{5}$$

Relation (5) implies the invariance of $D(V_\eta)$ under the unitaries u_g, $g\in G$. Since u acts irreducibly, $D(V_\eta)$ is dense in $\mathscr{H}$. It follows from Schur's Lemma (see Ref. 14, Chap. 2) that V_η is a multiple of an isometry. In particular, $D(V_\eta)=\mathscr{H}$ and V_η is bounded.

(c) Denote by p the atomic projection onto the subspace $\mathbb{C}\cdot\eta$,

$$\begin{aligned}
&\int_G\langle\xi|\alpha_s(p)\xi\rangle ds\\
&\quad=\int_G\langle\xi|u_spu_s^*\xi\rangle ds\\
&\quad=\int_G\langle u_s^*\xi|\eta\langle u_s^*\xi|\eta\rangle\rangle ds=\int_G|\langle u_s\eta|\xi\rangle|^2\,ds\\
&\quad=\|V_\eta\xi\|^2\leqslant\|V_\eta\|^2\|\xi\|^2\,,\quad\xi\in\mathscr{H}\,.
\end{aligned}\tag{6}$$

Therefore the operators $\int_K\alpha_s(p)ds$ are bounded by $\|V_\eta\|^2\cdot 1$ for arbitrary compact subsets $K\leqslant G$ and

$$\int_G\alpha_s(p)ds=\sup_{\substack{K\leqslant G\\ K\text{ compact}}}\int_K\alpha_s(p)ds$$

exists.

$\Leftarrow$: Suppose there is a nonvanishing operator $y\in\mathscr{B}(\mathscr{H})$ such that $\int_G\alpha_s(y^*y)ds<\infty$. Using spectral theory, a multiple $x\neq 0$ of an atomic projection p can be shown to exist such that

$$x\leqslant y^*y\Rightarrow\int_G\alpha_s(x)ds\leqslant\int_G\alpha_s(y^*y)ds<\infty$$

$$\Rightarrow\int_G\alpha_s(p)ds<\infty\,.$$

Relation (6) in (c) then implies that $\int_G|\langle u_s\eta|\xi\rangle|^2\,ds<\infty$, where η is a unit vector in $\mathscr{H}$ such that $p\mathscr{H}=\mathbb{C}\cdot\eta$ and $\xi\in\mathscr{H}$ is arbitrary. Q.E.D.

The above proof contains even more: One gets the following theorem, which slightly generalizes a well-known result on square-integrable (proper) representations of locally compact groups (see Ref. 13, 14.1, and Ref. 14, Theorem 2).

Theorem 3: Let u: $g\to u_g$, $g\in G$, be a square-integrable irreducible ray representation on the Hilbert space $\mathscr{H}$ with associated multiplier c. Then the representation u is unitarily equivalent to a subrepresentation of the left regular c-representation λ^c of G.

Proof: The operator V_η introduced in part (b) of the proof of Theorem 1 defines (by appropriate normalization) the isometry that, through relation (2), implements the equivalence between the unitary representation u and $r\lambda^cr$, where r is given as the projection with range $V_\eta\mathscr{H}$. Note that $V_\eta\mathscr{H}$ is λ^c-invariant by (2). [Remark: Theorem 2 can also be derived from Ref. 5, the proof of Lemma III.5.]

Of course the converse of theorem 2 is also true. It is formulated in Theorem 4.

Theorem 4: Let c: $G\times G\to\mathscr{T}$ be a multiplier and consider an irreducible ray representation u of the locally compact group G on a Hilbert space $\mathscr{H}$, which is unitarily equivalent to a subrepresentation of the left regular c-representation λ^c. Then u is square integrable.

Proof: Let r be a λ^c-invariant projection in $\mathscr{B}(L^2(G))$ such that u and $r\lambda^cr$ are unitarily equivalent and denote by M: $m\to M_m$, $m\in L_\infty(G)$, the multiplication representation of $L_\infty(G)$ on $L^2(G)$. Then $m\to rM_mr$, $m\in L_\infty(G)$, is a normal covariant embedding of $L_\infty(G)$ into

$$\left(r\mathscr{B}(L^2(G))r,G,\left\{\tilde{\alpha}_g(\cdot)\overset{\text{def}}{=}r\lambda^c(g)r\cdot r\lambda^c(g)^*r|g\in G\right\}\right)$$

$$\cong\left(\mathscr{B}(\mathscr{H}),G,\left\{\alpha_g(\cdot)\overset{\text{def}}{=}u_g\cdot u_g^*|g\in G\right\}\right)$$

and therefore $(\mathscr{B}(\mathscr{H}),G,\alpha)$ is integrable (see Ref. 5, Lemma

II.2). Thus u is square integrable (Theorem 1). Q.E.D.

Rephrased, the above theorems say that a unitary group representation u with multiplier c admits coherent states $|g\rangle$, $g\in G$, with $\int_G |g\rangle\langle g|dg < \infty$ if and only if u is a subrepresentation of the left regular c-representation. If the group G is compact, this is always the case; if G is noncompact, this condition is fulfilled only in very particular situations. Irreducible ray representations of the Galilei group, for example, never do admit coherent states in this sense.

V. PHASE SPACE REPRESENTATIONS ASSOCIATED WITH COHERENT STATES

In this chapter $u: g\to u_g$, $g\in G$, will denote a square integrable irreducible ray representation on a Hilbert space $\mathscr{H}$ with associated multiplier c, and $(\mathscr{B}(\mathscr{H}),G,\{\alpha_g = u_g \cdot u_g^* | g\in G\})$ will be the integrable ergodic W^*-system implemented by u.

Consider a normal covariant embedding χ: $L_\infty(G) \to \mathscr{B}(\mathscr{H})$ of $L_\infty(G)$ into $(\mathscr{B}(\mathscr{H}),G,\alpha)$. Here χ is a positive linear map and therefore completely positive since $L_\infty(G)$ is a commutative algebra (see Ref. 15, Theorem 4). It is therefore possible to use and to generalize Stinespring's theorem (see Ref. 15, Theorem 1). Accordingly, there exists a Hilbert space $\mathscr{K}$, an isometry $V: \mathscr{H}\to\mathscr{K}$, a faithful *-representation $\pi: L_\infty(G)\to\mathscr{B}(\mathscr{K})$ and a unitary ray representation $w: g\to w_g\in\mathscr{B}(\mathscr{K})$, $g\in G$ (whose associated multiplier is the same as that one of u), such that

$$
\begin{aligned}
&\text{(i) the linear subspace generated by } \pi(L_\infty(G))V\mathscr{H} \text{ is dense in } \mathscr{K},\\
&\text{(ii) } \chi(m) = V^*\pi(m)V,\ m\in L_\infty(G),\\
&\text{(iii) } w_g V = V u_g,\ g\in G,\\
&\text{(iv) } w_g \pi(m) w_g^* = \pi(\operatorname{Ad}\lambda(g)m),\ m\in L_\infty(G).
\end{aligned}
\tag{7}
$$

Proof: Considering the proof of Ref. 5, Theorem III.2, there is just one point that is not obvious at first sight, namely that $g\to w_g$, $g\in G$, is Borel: Every $\xi\in\mathscr{K}$ is approximated by a norm-bounded sequence $(\xi_n)_{n\in\mathbb{N}}$ of vectors that are linear combinations of vectors in $\pi(L_\infty(G))V\mathscr{H}$. For every $g\in G$ one has

$$
\begin{aligned}
&|\langle\xi|w_g\xi\rangle - \langle\xi_n|w_g\xi_n\rangle|\\
&\quad = \tfrac{1}{2}|\langle\xi-\xi_n|w_g(\xi+\xi_n)\rangle + \langle\xi+\xi_n|w_g(\xi-\xi_n)\rangle|\\
&\quad \leqslant \|\xi-\xi_n\|\,\|w_g\|\,\|\xi+\xi_n\|\\
&\quad \leqslant \|\xi-\xi_n\|(\|\xi\|+\|\xi_n\|) \underset{n\to\infty}{\longrightarrow} 0.
\end{aligned}
$$

Since the pointwise limit of Borel functions is Borel (see Ref. 16, 2.1.4) and since a linear combination of Borel functions is Borel, it is therefore sufficient to show that $\{g\to\langle\pi(m)V\xi|w_g\pi(n)V\eta\rangle_{\mathscr{K}}, g\in G\}$ is a Borel function for arbitrary $m,n\in L_\infty(G)$, $\xi,\eta\in\mathscr{H}$. Using the properties (7) (ii)–(7) (iv) one gets

$$
\begin{aligned}
&\{g\to\langle\pi(m)V\xi|w_g\pi(n)V\eta\rangle_{\mathscr{K}}\}\\
&\quad = \{g\to\langle\chi(m\operatorname{Ad}\lambda(g)n)^*\xi|u_g\eta\rangle_{\mathscr{H}}\}.
\end{aligned}
$$

That the latter function is Borel follows from the separability of $\mathscr{K}$ and the assumptions by use of Ref. 16, E.2 and E.9. Q.E.D.

If χ is extremal within the set of all normal covariant embeddings of $L_\infty(G)$ into $(\mathscr{B}(\mathscr{H}),G,\alpha)$, then $\pi(L_\infty(G))' = \pi(L_\infty(G))$ (see Ref. 5, Theorem III.2). This is now just the case if $\chi = \varphi_x$, where x is a multiple of an atomic projection with $\int_G \alpha_g(x)dg = 1$ (Lemma 2). Using Ref. 17, 2.8.3 and Ref. 18, III.1.4, one infers from the maximal commutativity of $\pi(L_\infty(G))$ in $\mathscr{B}(\mathscr{K})$ that π can be supposed to be the multiplication representation

$$
\begin{aligned}
&m\to M_m\in\mathscr{B}(L^2(G)),\quad m\in L_\infty(G),\\
&(M_m\Psi)(g) \overset{\text{def}}{=} m(g)\Psi(g),\quad g\in G,\quad \Psi\in L^2(G),
\end{aligned}
\tag{8}
$$

of $L_\infty(G)$ on the Hilbert space $L^2(G)$. Furthermore Ref. 19, Theorem 6 implies that $g\to w_g$, $g\in G$, can be assumed to be the left regular c-representation λ^c (cf. Theorem 7). This reproves Theorem 3 of Sec. IV, and gives in addition a structural result concerning normal covariant embeddings.

Theorem 5: Let u: $G\to\mathscr{B}(\mathscr{H})$ be a square-integrable irreducible representation, x be a multiple of an atomic projection with $\int_G \alpha_g(x)dg = 1$, and $\varphi_x: L_\infty(G)\to\mathscr{B}(\mathscr{H})$ the associated normal covariant embedding. Then there exists an isometry $V: \mathscr{H}\to\mathscr{B}(L^2(G))$ such that

(i) $\lambda^c(g)V = Vu_g$, $g\in G$,

(ii) $V^*M_mV = \varphi_x(m)$, $m\in L_\infty(G)$.

Remark: The proof given above puts φ_x: $L_\infty(G) \to\mathscr{B}(\mathscr{H})$ into the context of arbitrary normal embeddings. The fact that φ_x is extremal (Lemma 2) then leads to the particular structure given in Theorem 5. Theorem 5(ii) could also be shown to hold using the techniques of the proof of Theorem 2. The isometry V is just given as the (appropriately normalized) operator V_η introduced there.

In the sequel, $\beta_g(y) \overset{\text{def}}{=} \lambda^c(g)y\lambda^c(g)^*$, $y\in\mathscr{B}(L^2(G))$, $g\in G$, will denote the action on $\mathscr{B}(L^2(G))$ implemented by the left regular c-representation. The fixed point algebra $\mathscr{B}(L^2(G))^\beta$ is defined as

$$
\mathscr{B}(L^2(G))^\beta \overset{\text{def}}{=} \{y\in\mathscr{B}(L^2(G))|\,\beta_g(y)=y,\ \forall g\in G\}.
$$

Theorem 6: Let $c(r)$ denote the central support in $\mathscr{B}(L^2(G))^\beta$ of the projection $r \overset{\text{def}}{=} VV^*\in\mathscr{B}(L^2(G))^\beta$. Then there exists a unique (surjective σ-weakly continuous) *-isomorphism J: $\mathscr{B}(\mathscr{H})\to\{c(r)(\mathscr{B}(L^2(G))^\beta)'c(r)\}$ with the property

$$
V^*J(y)V=y,\quad y\in\mathscr{B}(\mathscr{H}).
$$

Furthermore, J is covariant, i.e., $\beta_g(J(y)) = J(\alpha_g(y))$, $g\in G$, holds for each element $y\in\mathscr{B}(\mathscr{H})$.

Proof: The subspace $c(r)L^2(G)$ is generated by $\{c(r)\mathscr{B}(L^2(G))^\beta c(r)\}V\mathscr{H}$ (see Ref. 20, 3.9); therefore Ref. 21, 1.3.1 implies that to each $y\in\mathscr{B}(\mathscr{H})$ there exists a unique operator $J(y)\in\{c(r)(\mathscr{B}(L^2(G))^\beta)'c(r)\}$ with the property $V^*J(y)V = y$. Here

$$
J:\ \mathscr{B}(\mathscr{H})\to\{c(r)(\mathscr{B}(L^2(G))^\beta)'c(r)\}
$$

is a σ-weakly continuous surjective *-isomorphism. Covariance follows from the uniqueness property by using that

$$\beta_g(J(y)) \in \{c(r)(\mathscr{B}(L^2(G))^\beta)'c(r)\}$$

and

$$V^*\beta_g(J(y))V$$
$$= V^*\lambda^c(g)J(y)\lambda^c(g)^*V$$
$$= u_g V^*J(y)Vu_g^* = u_g y u_g^* = \alpha_g(y), \quad y\in\mathscr{B}(\mathscr{H}).$$

Q.E.D.

Similarly as λ^c, a unitary ray representation ρ^c is defined on the Hilbert space $L^2(G)$,

$$(\rho^c(g)\xi)(s) \stackrel{\text{def}}{=} \Delta(g)^{1/2}c(sg,g^{-1})\xi(sg), \quad \xi\in L^2(G), \quad g\in G.$$

Here $\lambda^c(g_0)$ and $\rho^c(g)$ commute for arbitrary elements g_0, $g\in G$, i.e., $\rho^c(g)\in\mathscr{B}(L^2(G))^\beta$, $\forall g\in G$. The multiplier of λ^c is c itself. The associated multiplier τ of ρ^c is given by

$$\tau(g_1,g_2) = c(g_2^{-1},g_1^{-1}), \quad g_1,g_2\in G.$$

Theorem 7: The fixed point algebra $\mathscr{B}(L^2(G))^\beta$ is generated by the unitaries $\{\rho^c(g)|g\in G\}$. Its commutant $\{\mathscr{B}(L^2(G))^\beta\}'$ is generated by the left regular c-representation $\{\lambda^c(g)|g\in G\}$.

Proof: The duality theory of W^*-systems[22–24] is used; $\mathscr{R}(G)$ denotes the W^*-algebra generated by the right regular representation $\{\rho(g)|g\in G\}$. The maximal commutative subalgebra $\{M_m|m\in L_\infty(G)\}\subseteq\mathscr{B}(L^2(G))$ implements a coaction δ: $\mathscr{B}(L^2(G))\mapsto\mathscr{B}(L^2(G))\bar{\otimes}\mathscr{R}(G)$ (see Ref. 22, pp. 25–27 and Ref. 23, p. 1438) with the properties

(i) $\delta(\rho(g)) = \rho(g)\otimes\rho(g), \quad g\in G,$

(ii) $\delta(M_m) = M_m\otimes 1, \quad m\in L_\infty(G),$

(iii) $\delta(y) = y\otimes 1$ implies $y = M_m$ for suitable $m\in L_\infty(G)$,

(iv) $\delta(\mathscr{B}(L^2(G))^\beta)\subseteq\mathscr{B}(L^2(G))^\beta \,\bar{\otimes}\, \mathscr{R}(G)$.

Therefore $\delta(\rho^c(g)) = \rho^c(g)\otimes\rho(g), g\in G$, holds, and since

$$(\mathscr{B}(L^2(G))^\beta)^\delta \stackrel{\text{def}}{=} \{y\in\mathscr{B}(L^2(G))^\beta|\delta(y) = y\otimes 1\}$$
$$= \mathscr{B}(L^2(G))^\beta\cap\{M_m|m\in L_\infty(G)\}$$
$$= \mathbb{C}\cdot 1,$$

the first assertion of the theorem is now immediate from Ref. 19, Lemma 7. The second assertion is proved in a similar way. Q.E.D.

Since α is an ergodic representation, r is an atomic projection in $\mathscr{B}(L^2(G))^\beta$ and consequently, $c(r)$ is an atomic projection in the center $\mathscr{Z}(\mathscr{B}(L^2(G))^\beta)$ of $\mathscr{B}(L^2(G))^\beta$. *Thus there is a correspondence (not necessarily one-to-one) between square-integrable irreducible ray representations of a fixed multiplier c and the atomic projections in the center of the W^*-algebra generated by the left regular c-representation* (cf. Ref. 13, Chap. 15).

If $\mathscr{B}(L^2(G))^\beta$ is a factor, the only atomic projection in the center of $\{\mathscr{B}(L^2(G))^\beta\}'$ is the unit operator 1. In this situation, the covariant *-isomorphism J: $\mathscr{B}(\mathscr{H})\to\{\mathscr{B}(L^2(G))^\beta\}'$ is *unital* and the W^*-system

$$(\mathscr{B}(\mathscr{H}),G,\alpha)\cong(\{\mathscr{B}(L^2(G))^\beta\}',G,\beta)$$

is thus represented on the Hilbert space $L^2(G)$. This is, for example, the case if G is an *Abelian* group and α is a faithful representation (see Ref. 5, Lemma III.3). Incidentally, a square integrable irreducible ray representation u of an Abelian group G with associated multiplier c and fulfilling that α is faithful exists if and only if the group homomorphism s: $G\to\hat{G}$ defined by

$$c(t_0,t)^*\cdot c(t,t_0) = \langle s(t_0),t\rangle, \quad t_0,t\in G,$$

is an isomorphism (in particular surjective) (see Ref. 5, Theorem III.7). Here $\langle\cdot,\cdot\rangle$ describes the duality between G and its dual group $\hat{G}$.

The most simple example is given by $G = \mathbb{R}^{2n}$, $n = 1,2,\ldots$, and

$$c_\lambda((\mathbf{a}_1,\mathbf{b}_1),(\mathbf{a}_2,\mathbf{b}_2)) \stackrel{\text{def}}{=} \exp\{i\lambda(\mathbf{a}_1\mathbf{b}_2 - \mathbf{a}_2\mathbf{b}_1)\}, \quad \mathbf{a}_i,\mathbf{b}_i\in\mathbb{R}^n, \quad i = 1,2.$$

Here λ is a fixed positive real number corresponding to Planck's constant. Since $\mathbb{R}^{2n}$ can also be regarded as a phase space in the sense of classical mechanics, the representation J on $L^2(\mathbb{R}^{2n})$ is then referred to as a *phase space representation* of the Weyl relations. Theorem 6 can be regarded as a generalization of this phase space representation formalism, where $\mathbb{R}^{2n}$ is replaced by an arbitrary locally compact kinematical group.

Apart from the trivial case $G = \{e\}$, $\mathscr{H} = \mathbb{C}$, the representation J is highly reducible. Here $J(\mathscr{B}(\mathscr{H})) = (\mathscr{B}(L^2(G))^\beta)'$ and its commutant $\mathscr{B}(L^2(G))^\beta$ are (anti-)isomorphic, i.e., "equally large." A closer look shows that J is just the *standard representation* of $\mathscr{B}(\mathscr{H})$ (cf. Ref. 25, Chap. 2.5.4). The standard representation J: $\mathscr{B}(\mathscr{H})\to\mathscr{B}(L^2(G))$ has the nice property that every state ω on $\mathscr{B}(\mathscr{H})$ can be expressed by a vector $\xi\in L^2(G)$, $\omega(y) = \langle\xi|J(y)\xi\rangle, y\in\mathscr{B}(\mathscr{H})$.

Note that not only the respective quantum theories of a certain kinematical group G [with $\mathscr{B}(\mathscr{H})$ as "algebra of observables"] are represented on the Hilbert space $L^2(G)$, but also the respective classical theories: The algebra of observables is then given as the commutative algebra $L_\infty(G)$ [or $L_\infty(G/H)\subseteq L_\infty(G)$, where H is a closed subgroup of G], which is represented on $L^2(G)$ by the multiplication representation (7) ("Koopman-formalism"). This may considerably simplify the investigation of limits (such as $\hbar\to 0$) where a quantum theory goes into a classical theory. More generally, the emergence of classical observables could be discussed in such a framework.

APPENDIX: PROOF OF LEMMA 1

$\Rightarrow$: Let x be a positive operator from $\mathscr{M}$ such that $\int_G \alpha_g(x)dg = 1$ and let $\varphi_x(m) = \int_G \alpha_g(x)m(g)dg$, $m\in L_\infty(G)$, be the associated normal covariant embedding. Then the desired inequality follows from

$$\left\| \int_G \alpha_g(x) f(g) dg \right\|$$

$$= \sup_{\substack{\Psi \in \mathscr{M}_* \\ \|\Psi\| \leq 1}} \left| \left\langle \Psi, \int_G \alpha_g(x) f(g) dg \right\rangle \right|$$

$$= \sup_{\|\Psi\| \leq 1} \left| \int_G \Psi(\alpha_g(x)) f(g) dg \right|$$

$$\leqslant \sup_{\|\Psi\| \leq 1} \|\Psi\| \cdot \|x\| \cdot \int_G f(g) dg$$

$$= \|x\| \cdot \int_G f(g) dg, \quad \forall f \in \{L^1(G) \cap L_\infty(G)\}_+ .$$

$\Leftarrow$: Consider a fundamental system $(K_n)_{n\in\mathbb{N}}$ of compact neighborhoods of the neutral element $e\in G$ and continuous positive functions $(k_n)_{n\in\mathbb{N}}$ with the properties

(i) $\int_G k_n(g)\Delta(g^{-1})dg = 1$,

(ii) support $(k_n) \subseteq K_n$, $n\in\mathbb{N}$

(Δ is the modular function of G).

It will be shown first that the set of operators $(\chi(k_n))_{n\in\mathbb{N}}$ is norm-bounded: Since Δ: $G \rightarrow \mathbb{R}^+$ is continuous, one can find an $N_\epsilon \in \mathbb{N}$ such that $|\Delta(g) - 1| < \epsilon$ holds for all $g\in K_n$, $n \geqslant N_\epsilon$:

$$\Rightarrow \left| \int_G k_n(g) dg - \int_G k_n(g)\Delta(g^{-1}) dg \right|$$

$$= \left| \int_G k_n(g)\Delta(g^{-1})(\Delta(g) - 1) dg \right|$$

$$\leqslant \int_G k_n(g)\Delta(g^{-1})|\Delta(g) - 1| dg$$

$$\leqslant \left\{ \int_G k_n(g)\Delta(g^{-1}) dg \right\} \cdot \epsilon, \quad n \geqslant N_\epsilon$$

$$\Rightarrow \|\chi(k_n)\| \leqslant N \cdot \int_G k_n(g) dg \leqslant N(1+\epsilon), \quad \forall n \geqslant N_\epsilon$$

$\Rightarrow (\|\chi(k_n)\|)_{n\in\mathbb{N}}$ is bounded by a positive number $d\in\mathbb{R}_+$.

Since $\mathscr{M}_d \stackrel{\text{def}}{=} \{y \in \mathscr{M} \mid \|y\| \leqslant d\}$ is σ-weakly compact and second countable, there exists a subsequence $(\chi(k_{n(j)}))_{j\in\mathbb{N}}$ converging σ-weakly to a positive operator $x\in\mathscr{M}_d$. Due to Ref. 5, Proof of Lemma II.2, and Ref. 25, Theorem 2.7.11, one has $\int_G \alpha_g(x)dg \leqslant 1$. For $f\in\{L^1(G)\cap L_\infty(G)\}$ and arbitrary $\Psi\in\mathscr{M}_*$, the following holds:

$$\left| \left\langle \Psi, \int_G \alpha_g(x) f(g) dg \right\rangle - \left\langle \Psi, \int_G \alpha_g(\chi(k_{n(j)})) f(g) dg \right\rangle \right|$$

$$= \left| \int_G \{\Psi(\alpha_g(x)) - \Psi(\alpha_g(\chi(k_{n(j)})))\} f(g) dg \right|$$

$$\leqslant \int_G |\Psi(\alpha_g(x)) - \Psi(\alpha_g(\chi(k_{n(j)})))| \, |f(g)| dg \xrightarrow[j\to\infty]{} 0 .$$

In the last step Lebesgue's dominated convergence theorem has been used. Thus $\varphi_{\chi(k_{n(j)})}(f)$ converges σ-weakly to $\varphi_x(f)$ for all $f\in\{L^1(G)\cap L_\infty(G)\}$. In the proof of Ref. 5, Theorem II.3 it is shown that $\varphi_{\chi(k_n)}(m)$ converges σ-weakly to $\chi(m)$ for $m\in L_\infty(G)$ if $n\rightarrow\infty$. The same is true for the subsequence $n(j)$, $j\in\mathbb{N}$. Therefore $\chi(f) = \varphi_x(f)$, $\forall f\in\{L^1(G)\cap L_\infty(G)\}$, holds true.

$\{L^1(G)\cap L_\infty(G)\}$ is σ-weakly dense in $L_\infty(G)$. If m is an arbitrary element of $L_\infty(G)$, approximated σ-weakly by a net $(f_\beta)_{\beta\in I}, f_\beta\in\{L^1(G)\cap L_\infty(G)\}, \beta\in I$ (where I is an index set), it follows from the σ-weak continuity of φ_x and χ that

$$\varphi_x(m) = \lim_\beta \varphi_x(f_\beta) = \lim_\beta \chi(f_\beta) = \chi(m)$$

$$\Rightarrow \varphi_x(m) = \chi(m), \quad \forall m\in L_\infty(G) . \qquad \text{Q.E.D.}$$

[1] J. R. Klauder and B. S. Skagerstam, *Coherent States. Applications in Physics and Mathematical Physics* (World Scientific, Singapore, 1984).

[2] A. M. Perelomov, Commun. Math. Phys. **26**, 222 (1972).

[3] S. K. Berberian, *Lectures in Functional Analysis and Operator Theory* (Springer, New York, 1974).

[4] H. Primas, *Chemistry, Quantum Mechanics and Reductionism. Perspectives in Theoretical Chemistry* (Springer, Berlin, 1983), 2nd corrected edition.

[5] A. Amann, "Observables in W^*-algebraic quantum mechanics," to appear in Fortschr. Phys.

[6] H. Scutaru, Lett. Math. Phys. **2**, 101 (1977).

[7] H. Moscovici and A. Verona, Ann. Inst. H. Poincaré **29**, 139 (1978).

[8] D. DeSchreye and H. H. Zettl, "Notes on integrable W^*-dynamical systems," preprint, Kath. Universiteit Leuven, 1983.

[9] A. Connes and M. Takesaki, Tôhôku Math. J. **29**, 473 (1977).

[10] A. Amann, thesis ETH no. 7517, Zürich, 1984.

[11] E. B. Davies, *Quantum Theory of Open Systems* (Academic, London, 1976).

[12] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).

[13] J. Dixmier, *C*-Algebras* (North-Holland, Amsterdam, 1977).

[14] M. Duflo and C. C. Moore, J. Funct. Anal. **21**, 209 (1976).

[15] W. F. Stinespring, Proc. Am. Math. Soc. **6**, 211 (1955).

[16] D. L. Cohn, *Measure Theory* (Birkhäuser, Boston, 1980).

[17] G. K. Pedersen, *C*-Algebras and Their Automorphism Groups* (Academic, London, 1979).

[18] M. Takesaki, *Theory of Operator Algebras* (Springer, New York, 1979), Vol. I.

[19] Y. Nakagami and C. E. Sutherland, Pac. J. Math. **83**, 221 (1979).

[20] S. Stratila and L. Zsido, *Lectures on von Neumann Algebras* (Editura Academiei, Bucuresti and Abacus Press, Tunbridge Wells, 1975).

[21] W. B. Arveson, Acta Math. **123**, 142 (1969).

[22] Y. Nakagami and M. Takesaki, *Duality for Crossed Products of von Neumann Algebras, Lecture Notes in Mathematics*, Vol. 731 (Springer, Berlin, 1979).

[23] S. Stratila, D. Voiculescu, and L. Zsido, Rev. Roumaine Math. Pures Appl. **21**, 1411 (1976).

[24] S. Stratila, D. Voiculescu, and L. Zsido, Rev. Roumaine Math. Pures Appl. **22**, 83 (1977).

[25] O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics* (Springer, New York, 1979), Vol. 1.

Evolution loops

Bogdan Mielnik[a)]
Departamento de Física, CINVESTAV, Apdo. Postal 14-740, 07000, México, D. F., Mexico

(Received 2 December 1985; accepted for publication 5 February 1986)

The problem of manipulating Schrödinger's particle by time-dependent external fields is discussed. New solutions of the evolution problem, called *evolution loops*, are found. They correspond to the "particle memory" in the sense of Brewer and Hahn [Sci. Am. **251**(12), 50 (1984)]. A technique of generating the unitary operations by perturbing the evolution loops is outlined.

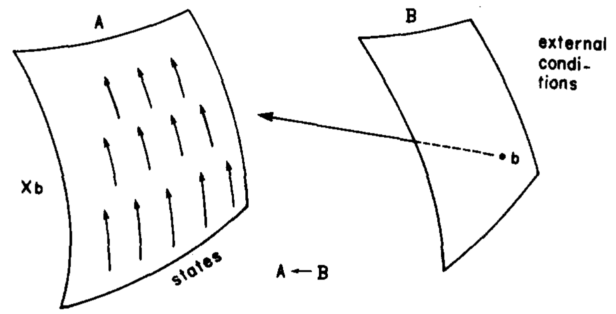
I. INTRODUCTION

There is a certain asymmetry in the present-day dynamic theories, which seem to favor the theory of closed systems. When speaking about nondissipative dynamics, one usually has in mind a system with a fixed Hamiltonian. Its evolution is represented by a given one-parameter group of transformations (time translations). In cases of extreme simplification, an image of the dynamics is obtained just by considering discrete iterations of one single mapping.[1-3] In these pictures, the data representing the external world are fixed and the effort of the theory is centered around resolving the actual evolution problem. The question about the comparative behavior in alternative universes is left untouched. Yet, if the physical theories were at all created, it is only because the physical systems are open. The real universe is a place where the external conditions can be changed and experiments can be performed. An idea thus arises that a nontrivial dynamic theory should not deal with a one-parameter transformation group, but with a wider family of transformations that are available to the experimenter.[4-9]

II. THE PROBLEM OF MANIPULATION

The most obvious picture of an open, dissipation free system is obtained by postulating the existence of a certain *manifold of pure states* A and a certain variety B of the *external conditions* available, and by assuming that the external conditions $b\in B$ generate some "tendencies to evolve" on A (see Ref. 8).

Definition 1: A simple open system is a pair of generalized differential manifolds A and B, of dimension finite or not, called the manifold of states and the manifold of the external conditions together with a mapping $b\to X_n$, which to each $b\in B$ assigns an integrable vector field X_b on A. The field X_b is interpreted as the evolution law for the states $a\in A$ under the influence of the external condition $b\in B$. In what follows, we shall say that B controls A $(B\to A)$ or that B is a manifold of generators for A:



(The above concept of a dynamic system is close, though not identical, to that in terms of categories.[10])

Definition 1 suggests a certain new question in a dynamic theory. In the traditional approach to dynamics one *knows* the external conditions and one asks what will be the motion of the system? The question we would like to ask is the inverse one. Given a certain transformation $g: A\to A$ in a space of physically interesting states, can this transformation be accomplished dynamically? If so, which method should be applied? These questions correspond to a "Rubik's cube view" of the dynamic system as opposed to the "one-parameter scheme," see also Refs. 5, 8, and 11.

Below, we shall take merely first steps in collecting some operational answers. They will concern the evolution processes of a nonrelativistic particle in time-dependent external fields. Our models will not depart from the Schrödinger wave mechanics. Yet, they have been investigated little in the past, because of the difficulty of applying the Baker–Campbell–Hausdorff formula. Hence, it might be of interest that the problem of the evolution of Schrödinger's particle in time-dependent external fields also has its exact solutions. They turn out to be a natural starting point for convenient manipulation procedures. All the results presented below are purely heuristic. They are obtained by following Schrödinger's quantum mechanics with variable fields "to the letter." The problem of the applicability limits, as well as the questions of essential self-adjointness and strong convergence, are open.

III. THE SIMPLEST EVOLUTION LOOPS (ONE SPACE DIMENSION)

What kind of evolution operations of the Schrödinger's

[a)] On leave of absence from the Institute of Theoretical Physics, Warsaw University, Warsaw, Poland.

wave packet can be induced by an arbitrary time-dependent potential $V(q,t)$? The corresponding differential equation

$$\frac{d}{dt}U(t,t_0) = -iH(t)U(t,t_0), \quad U(t_0,t_0) = 1, \tag{3.1}$$

with $H(t) = \frac{1}{2}p^2 + V(q,t)$, $q = x$, $p = (1/i)(d/dx)$ (put $m = \hbar = 1$), has been traditionally treated in the framework of the formalism of the "chronological products" (see Ref. 12). Yet, due to the recent development of algebraic and computer techniques,[13-17] some competitive methods become relevant. Below, we shall study $U(t,t_0)$ in light of the following simple lemma,[18] which states that any unitary U: $L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is determined up to a phase factor by the corresponding transformation of q and p: $q \to UqU^*$, $p \to UpU^*$.

Lemma 1: If U_1 and U_2 are two unitary operators in $L^2(\mathbb{R})$ and $U_1qU_1^* = U_2qU_2^*$, $U_1pU_1^* = U_2pU_2^*$, then $U_2 = e^{i\alpha}U_1 (\alpha \in \mathbb{R})$.

Proof:

$$U_1qU_1^* = U_2qU_2^* \Rightarrow qU_1^*U_2 = U_1^*U_2q$$

and

$$U_1pU_1^* = U_2pU_2^* \Rightarrow pU_1^*U_2 = U_1^*U_2p.$$

Since $L^2(\mathbb{R})$ is an irreducible representation space of the Heisenberg algebra, the operator that commutes with both q and p must be a number; hence, $U_1^*U_2 = e^{i\alpha}$ $(\alpha \in \mathbb{R})$.

An interesting subclass of exact solutions of (3.1) appears if $V(q,t)$ is permitted to oscillate fast. The simplest one of them is obtained by taking in (3.1) a strong external potential $(1/\epsilon)V(q)$ $(0 < \epsilon \sim 0)$ and by letting it act during a short time interval $[t_0, t_0 + \epsilon]$. The corresponding evolution operator is

$$\begin{aligned} U_\epsilon &= U(t_0, t_0 + \epsilon) \\ &= \exp\{-i\epsilon[p^2/2 + (1/\epsilon)V(q)]\} \\ &= \exp\{-i[\epsilon(p^2/2) + V(q)]\} \end{aligned} \tag{3.2}$$

and for $\epsilon \to 0$ it becomes

$$e^{-iV(q)}. \tag{3.3}$$

The operator (3.3) might be interpreted as the exact solution of the evolution problem (3.1) obtained for the δ-like pulse of the external potential:

$$V(q,t) = \delta(t - t_0)V(q). \tag{3.4}$$

Now, by considering the evolution processes generated by many different "shocks" separated by intervals of free evolution, one arrives at the following class of the evolution operators:

$$e^{-i\tau_1(p^2/2)}e^{-iV_1(q)} \cdots e^{-i\tau_n(p^2/2)}e^{-iV_n(q)}, \tag{3.5}$$

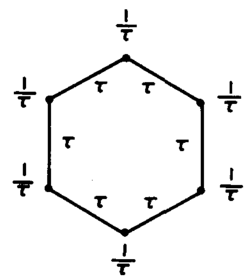
where $\tau_1, \ldots, \tau_n \geqslant 0$, $n = 1, 2, \ldots$. The operators (3.5) represent what might be called "the solutions of the evolution problem in pulsating fields." The mechanism of Trotter formulas suggests that they can approximate any $U(t,t_0)$. A question now arises: How rich is the class (3.5)? What evolution processes can be generated just by multiple potential pulses? It turns out that the answer depends on a special phenomenon that can occur for the operator sequences (3.5).

A. "The echo"

One of the seldom studied aspects of a dynamic theory is the existence of nontrivial sequences of dynamic events that "rotate the space of states around itself" yielding the identity transformation (see also Ref. 5). Any such sequence will be called an "evolution loop." One of most striking examples of this phenomena is the *spin echo*. We shall show that, in agreement with predictions of Brewer and Hahn,[19] the "echo" is not exclusive for the spin degrees but can also occur in $L^2(\mathbb{R})$. One of the proofs is due to the following simple identity[18]:

$$\underbrace{e^{-i\tau p^2/2}e^{-i(1/\tau)q^2/2} \cdots e^{-i\tau p^2/2}e^{-i(1/\tau)q^2/2}}_{12 \text{ terms}} = 1. \tag{3.6}$$

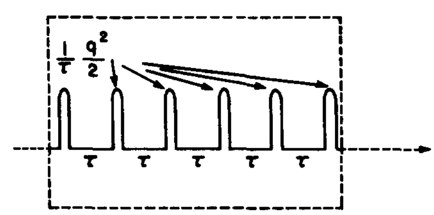
Note, that all signs here are identical. The formula (3.6) means that, after applying six pulses of the oscillator-shaped potential intertwined with six intervals of the free evolution, each nonrelativistic charged particle returns to its initial state. This can be visualized by the "evolution loop":



The numbers at the vertices symbolize the intensity of the oscillator-shaped pulses, while the sides correspond to the intervals of the free evolution. A consequence of (3.6) is the "inversion formula":

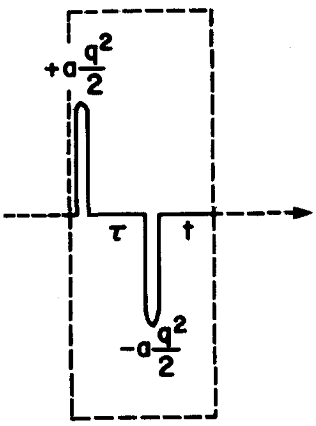
$$\underbrace{e^{-i(1/\tau)q^2/2}e^{-i\tau p^2/2} \cdots e^{-i\tau p^2/2}e^{-i(1/\tau)q^2/2}}_{11 \text{ terms}} = e^{i\tau p^2/2}. \tag{3.7}$$

This means that, under the influence of an "evolution sandwich" composed of six pulses of the oscillator potentional and five "rest intervals," all nonrelativistic particles are induced to "go back in time," by performing the operation $e^{i\tau p^2/2}$, which is inverse to the free evolution (see the diagram below). This provides an analog in $L^2(\mathbb{R})$ of the "rotating dye experiment" of Brewer and Hahn[19]:



It turns out that (3.7) is only one of many methods to achieve $e^{i\tau p^2/2}$. Some alternative prescriptions are worth studying. The most useful of them arise by considering sequences of opposite oscillator pulses, one attractive and one repulsive; the typical element of the process being

$$U = e^{-itp^2/2} e^{+iaq^2/2} e^{-i\tau p^2/2} e^{-iaq^2/2}, \tag{3.8}$$



The structure of this operation can be examined with the help of Lemma 1. In agreement with the identity

$$e^{\alpha A} B e^{-\alpha A} = B + \frac{\alpha}{1!}[A,B] + \frac{\alpha^2}{2!}[A,[A,B]] + \cdots, \tag{3.9}$$

the operators $e^{iaq^2/2}$ and $e^{-i\tau p^2/2}$ generate the following linear transformations of q and p:

$$e^{iaq^2/2}\begin{pmatrix} q \\ p \end{pmatrix} e^{-iaq^2/2} = \begin{pmatrix} 1 & 0 \\ -a & 1 \end{pmatrix}\begin{pmatrix} q \\ p \end{pmatrix}, \tag{3.10}$$

$$e^{-i\tau p^2/2}\begin{pmatrix} q \\ p \end{pmatrix} e^{i\tau p^2/2} = \begin{pmatrix} 1 & -\tau \\ 0 & 1 \end{pmatrix}\begin{pmatrix} q \\ p \end{pmatrix}. \tag{3.11}$$

Hence, the operator U in (3.8) corresponds to the transformation matrix

$$u = \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix}\begin{pmatrix} 1 & -\tau \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -a & 1 \end{pmatrix}\begin{pmatrix} 1 & -t \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 + \tau a, & -(t + \tau + at\tau) \\ \tau a^2, & 1 - a\tau - a^2 t\tau \end{pmatrix}. \tag{3.12}$$

In view of Lemma 1, the matrix u determines U with the accuracy up to a phase factor. The cases of special interest are $u^n = \mathbb{1}$ $(n = 1,2,\ldots)$. Then by Lemma 1, U^n must be proportional to 1. (We shall write $U^n \equiv 1$, without worrying about the phase factor.) Now, the identity $u^n = \mathbb{1}$ can hold if the roots of the characteristic polynomial

$$D_u(\lambda) = \lambda^2 - \lambda \operatorname{Tr} u + 1 = \lambda^2 - (2 - a^2 t\tau)\lambda + 1$$

$$= \lambda^2 - \kappa\lambda + 1, \tag{3.13}$$

are simultaneously roots of unity:

$$\lambda_{1,2} = (\kappa \mp \sqrt{\kappa^2 - 4})/2 \in \bigcup_n \sqrt[n]{1}, \quad \kappa = 2 - a^2 t\tau. \tag{3.14}$$

The matrix (3.12) is nondegenerate and the numbers $\lambda_{1,2}$ both have the absolute value 1, if $-2 < \kappa < 2 \Rightarrow a^2 t\tau < 4$. In addition, they become roots of unity for a sequence of values $\kappa = 0 - 1, 1, \ldots$; in general $\kappa_{nl} = 2\cos 2\pi l/n$ $(n \geqslant 3,$ $0 < 2l < n)$. Each of these values generates a certain "closed loop of the evolution operations." Some examples are for

$\kappa = 0, \quad \lambda_{1,2} = \pm i,$

$$[e^{-itp^2/2} e^{\mp i(2/t\tau)^{1/2} q^2/2} \times e^{-i\tau p^2/2} e^{\pm i(2/t\tau)^{1/2} q^2/2}]^4 \equiv 1; \tag{3.15}$$

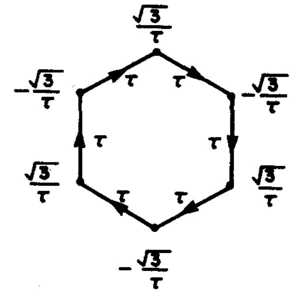
for $\kappa = 1, \quad \lambda_{1,2} = -\frac{1}{2} \pm i\sqrt{3}/2,$

$$[e^{-itp^2/2} e^{\mp i(3/t\tau)^{1/2} q^2/2} \times e^{-i\tau p^2/2} e^{\pm i(3/t\tau)^{1/2} q^2/2}]^3 \equiv 1; \tag{3.16}$$

and a general case, where $\kappa = 2\cos 2\pi l/n$,

$$[e^{-itp^2/2} e^{\mp i(2\sin(\pi l/n)/\sqrt{t\tau}) q^2/2} \times e^{-i\tau p^2/2} e^{\mp i(2\sin(\pi l/n)/\sqrt{t\tau}) q^2/2}]^n \equiv 1. \tag{3.17}$$

For $t = \tau$, the formula (3.16) defines an evolution loop that is a changing-sign alternative of that of the formula (3.6):



The sign-changing sequences of shocks are of special interest, as they provide source-free pulses in three dimensions. Some other cases of such loops are represented in Figs. 1 and 2.

Each of the loop formulas (3.15)–(3.18) yields simultaneously a method of inverting the free evolution by a sequence of sign-changing pulses of the $q^2/2$ potential. The simplest such prescription is derived from (3.16):

$$\underbrace{e^{\mp i(\sqrt{3}/t\tau)q^2/2} e^{-itp^2/2} e^{\pm i(\sqrt{3}/t\tau)q^2/2} e^{-i\tau p^2/2} \cdots}_{11 \text{ terms}} \equiv e^{i\tau p^2/2}, \tag{3.18}$$

while the general one is

$$e^{i(a/\sqrt{t\tau})q^2/2} e^{-itp^2/2} e^{-i(a/\sqrt{t\tau})q^2/2} e^{-i\tau p^2/2} \cdots e^{-i(a/\sqrt{t\tau})q^2/2} \equiv e^{i\tau p^2/2}$$

$$[4n - 1 \text{ terms}; \quad a = \pm 2\sin(\pi l/n)]. \tag{3.19}$$

What is special about the formulas (3.4) and (3.19) is that they are operator identities, i.e., kinds of universal prescriptions, which can be engineered "in blind." When ap-

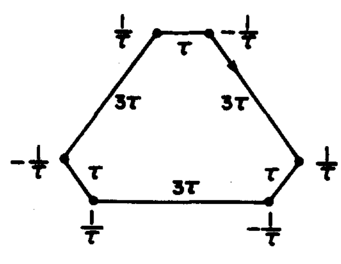


FIG. 1. One of the loops of the formula (3.16).

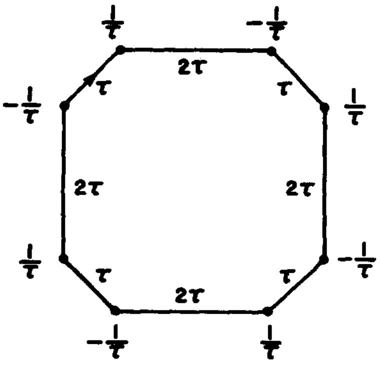


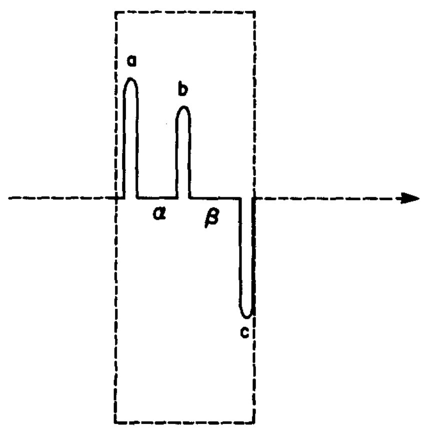
FIG. 2. A loop of formula (3.15) for $t = 2\tau$.

plied to a free Schrödinger's particle, they permit us to stop the free evolution and restore the particle state from the past without actually knowing what this state was and which is the present state of the particle (see also Refs. 20 and 21). Due to a simple mathematical mechanism[18] the existence of such effects in $L^2(\mathbf{R})$ is essential for the possibility of generating arbitrary unitary operations (see also Sec. V).

B. Question of "time economy"

The prescriptions quoted above suffer some "time disadvantage." The simplest method of achieving $e^{i\tau p^2/2}(\tau > 0)$ read from (3.16) requires six potential shocks and five intervals of the free evolution of the length $2\tau + 3t > 2\tau$. Hence, in order to "rejuvenate" the system for a time τ, one should have to work at least twice as long. Is that limitation necessary? To check this, consider the unitary operator corresponding to three arbitrary pulses of $q^2/2$ potential and two arbitrary intervals of the free evolution:

$$U_{\alpha ab\beta c} = e^{-icq^2/2} e^{-i\beta p^2/2} e^{-ibq^2/2} e^{-i\alpha p^2/2} e^{-iaq^2/2}, \tag{3.20}$$



The corresponding transformation matrix for q, p is

$$u = \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix} \begin{pmatrix} 1 & -\alpha \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix} \begin{pmatrix} 1 & -\beta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix} = \begin{pmatrix} 1 - \alpha b - \alpha c - \beta c - \alpha\beta bc, & -\alpha - \beta + \alpha\beta b \\ a + b + c - \alpha(b+c)a + , & 1 - \alpha a - \beta a + \\ - \beta(a+b)c + \alpha\beta abc & -\beta b + \alpha\beta ab \end{pmatrix}. \tag{3.21}$$

Demanding that

$$u = \begin{pmatrix} 1 & \tau \\ 0 & 1 \end{pmatrix}, \quad \tau \in \mathbf{R}, \tag{3.22}$$

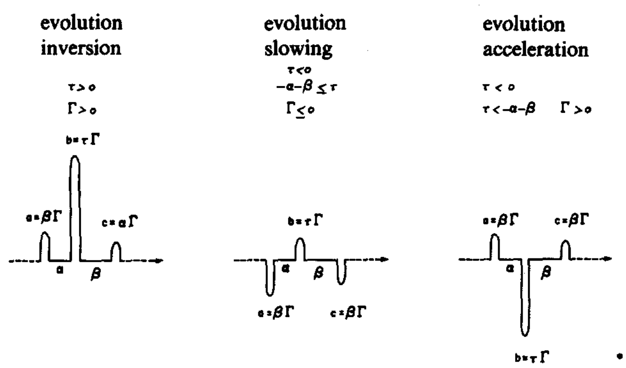
one obtains four equations for α, β, a, b, c, which yield

$$\begin{pmatrix} a = \beta\Gamma, \quad b = \tau\Gamma, \quad c = \alpha\Gamma, \\ \Gamma = (\alpha + \beta + \tau)/\alpha\beta\tau. \end{pmatrix} \tag{3.23}$$

If this holds, then by Lemma 1

$$e^{-i\alpha\Gamma q^2/2} e^{-i\beta p^2/2} e^{-i\tau\Gamma q^2/2} e^{-i\alpha p^2/2} e^{-i\beta\Gamma q^2/2} \equiv e^{i\tau p^2/2},$$
$$\Gamma = (\alpha + \beta + \tau)/\alpha\beta\tau. \tag{3.24}$$

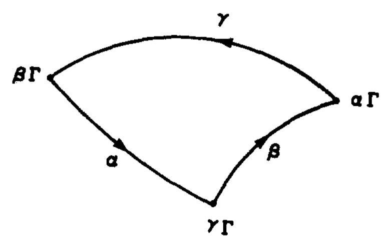
One thus sees that the operations $e^{i\tau p^2/2}$, with τ positive or negative, can be generated always at the expense of two arbitrarily short time intervals $\alpha, \beta > 0$. If $\tau > 0$, the formula (3.24) represents the "retrospection operation" achieved by three pulses of the attractive oscillator potential. If $-\alpha - \beta < \tau < 0$, (3.24) does not mean the inversion, but "slowing down" of the natural turn of the evolution, and is achieved by two negative and one positive pulse. If $\tau < -\alpha - \beta$, (3.24) yields the acceleration of the free evolution, engineered by two positive and one negative pulse:



Denoting $\tau = \gamma$, these three prescriptions are reduced to the "circular identity"

$$e^{-i\gamma p^2/2} e^{-i\alpha\Gamma q^2/2} e^{-i\beta p^2/2} e^{-i\gamma q^2/2} e^{-i\alpha p^2/2} e^{-i\beta\Gamma q^2/2} \equiv 1,$$
$$\Gamma = (\alpha + \beta + \gamma)/\alpha\beta\gamma, \tag{3.25}$$

visualized by the evolution loop:



Its existence implies[18] the following lemma.

Lemma 2: The dynamic operations associated with the pulsating fields can generate any unitary operation in $L^2(\mathbf{R})$ within an arbitrarily short time.

Our results, until now, are exact solutions of the evolution problem (3.1). However, Schrödinger's quantum mechanics is only an approximate scheme. It neglects the electrodynamic effects of quickly changing potentials, as well as the possibility of pair creation. Hence, our solutions are formal. A question arises: can the same effects be achieved by slowly changing fields? Can the system be induced to perform $e^{i\tau p^2/2}$ ($\tau > 0$) as a result of a "soft persuasion" instead of "brutal force"? As far as the quadratic potentials are considered, this leads to an old but unsolved problem.

IV. EVOLUTION IN TIME-DEPENDENT QUADRATIC POTENTIALS. CONTINUOUS LOOPS

How does Schrödinger's particle evolve in the presence of a time-dependent Hamiltonian $H(t) = p^2/2 + \gamma(t)q^2/2$? The corresponding evolution operator $U(t)$ obeys the differential equation

$$\frac{dU}{dt} = -i\left(\frac{p^2}{2} + \gamma(t)\frac{q^2}{2}\right)U(t), \quad U(0) = 1, \tag{4.1}$$

which can be viewed as a limiting case of the Baker–Campbell–Hausdorff composition problem for a sequence of exponentials $e^{-iA_1}e^{-iA_2}\cdots e^{-iA_n}$, where A_j ($j = 1,...,n$) are quadratic forms in q,p. One of the most elegant solutions of this last, discrete problem is due to the correspondence between the unitary operators e^{-iA} (with A quadratic in q,p) and the Lorentz rotations in the three-dimensional Minkowski space (Plebański).[22] A profound study of the quadratic exponents is found in Wolf.[23] The above discrete results allow us to predict that the solution of the continuous problem (4.1) should be of the form

$$U(t) = \exp\left(-i\left[a\frac{p^2}{2} + b\frac{pq+qp}{2} + c\frac{q^2}{2}\right]\right), \tag{4.2}$$

where $a = a(t)$, $b = b(t)$, and $c = c(t)$ are three real functions. However, the problem of finding (4.2) amounts to a continuous composition of the pseudo-Euclidean rotations.[24,25] One might thus think that Eq. (4.1) is not effectively solvable. Yet, this is not always the case.

A. Iterative solution

One of the algorithms used to solve Eq. (4.1) stems from Lemma 1. It assures that the operator $U(t)$ in (4.1) is defined by the corresponding transformation of q,p. Due to (4.2) this transformation is linear, and since it conserves the commutator $[q,p] = i$, it must be given by a certain unimodular matrix $u(t)$:

$$U(t)\begin{pmatrix} q \\ p \end{pmatrix}U(t)^* = u(t)\begin{pmatrix} q \\ p \end{pmatrix} = \begin{pmatrix} F(t) & -T(t) \\ P(t) & G(t) \end{pmatrix}\begin{pmatrix} q \\ p \end{pmatrix}, \tag{4.3}$$

with

$$F(t)G(t) + P(t)T(t) = \operatorname{Det} u(t) = 1. \tag{4.4}$$

Differentiating now both sides of (4.3) with respect to t and then using (4.1) and its conjugate, one sees

$$\begin{aligned}\frac{du(t)}{dt}\begin{pmatrix} q \\ p \end{pmatrix} &= -i\left[H(t),u(t)\begin{pmatrix} q \\ p \end{pmatrix}\right] \\ &= -iu(t)\left[\left(\frac{p^2}{2} + \gamma(t)\frac{q^2}{2}\right),\begin{pmatrix} q \\ p \end{pmatrix}\right] \\ &= u(t)\begin{pmatrix} -p \\ \gamma(t)q \end{pmatrix} = u(t)\Gamma(t)\begin{pmatrix} q \\ p \end{pmatrix}\end{aligned} \tag{4.5}$$

with

$$\Gamma(t) = \begin{pmatrix} 0 & -1 \\ \gamma(t) & 0 \end{pmatrix}. \tag{4.6}$$

Henceforth

$$\frac{du(t)}{dt} = u(t)\Gamma(t), \quad u(0) = 1, \tag{4.7}$$

or equivalently

$$u(t) = \mathbb{1} + \int_0^t u(t')\Gamma(t')dt'. \tag{4.8}$$

The iterative solution is

$$u(t) = \mathbb{1} + \int_0^t \Gamma(t_1)dt_1 + \int_0^t\int_0^{t_1} \Gamma(t_2)\Gamma(t_1)dt_2\,dt_1 + \int_0^t\int_0^{t_1}\int_0^{t_2} \Gamma(t_3)\Gamma(t_2)\Gamma(t_1)dt_3\,dt_2\,dt_1 + \cdots. \tag{4.9}$$

It yields

$$\begin{aligned}F(t) = 1 &- \int_0^t \gamma(t_1)t_1dt_1 + \int_0^t\int_0^{t_1} \gamma(t_1)(t_1 - t_2)\gamma(t_2)t_2\,dt_2\,dt_1 \\ &- \int_0^t\int_0^{t_1}\int_0^{t_2} \gamma(t_1)(t_1 - t_2)\gamma(t_2)(t_2 - t_3)\gamma(t_3)t_3\,dt_3\cdots dt_1 + \cdots,\end{aligned} \tag{4.10}$$

$$\begin{aligned}G(t) = 1 &- \int_0^t (t - t_1)\gamma(t_1)dt_1 + \int_0^t\int_0^{t_1} (t - t_1)\gamma(t_1)(t_1 - t_2)\gamma(t_2)dt_2\,dt_1 \\ &- \int_0^t\int_0^{t_1}\int_0^{t_2} (t - t_1)\gamma(t_1)(t_1 - t_2)\gamma(t_2)(t_2 - t_3)\gamma(t_3)dt_3\cdots dt_1 + \cdots,\end{aligned} \tag{4.11}$$

$$P(t) = -G'(t), \tag{4.12}$$

$$T(t) = \int_0^t F(t')dt'. \tag{4.13}$$

The trace invariant that decides about the algebraic properties of $u(t)$ henceforth is

$$\operatorname{Tr} u(t)=2-t\int_0^t \gamma(t_1)dt_1+\int_0^t\int_0^{t_1}(t-t_1+t_2)\gamma(t_1)(t_1-t_2)\gamma(t_2)dt_2\,dt_1$$
$$-\int_0^t\int_0^{t_1}\int_0^{t_2}(t-t_1+t_3)\gamma(t_1)(t_1-t_2)\gamma(\tau_2)(t_2-t_3)\gamma(t_3)dt_3\,dt_2\,dt_1+\cdots. \tag{4.14}$$

B. The loop criterion

Now, let g: $\mathbb{R}\to\mathbb{R}$ be a bounded, piecewise continuous, periodic function of period 1. Then choose as $\gamma(t)$ in (4.1) the function of period T: $\gamma(t)=ag(t/T)$. Can $\gamma(t)$ generate the loops in $[0,nT]$? Consider the evolution matrix $u(T)$. Its trace, due to (4.14), becomes

$$\operatorname{Tr} u(T)=2+\sum_{n=1}^{+\infty}R_n\alpha^n=R(\alpha),\quad \alpha=aT^2, \tag{4.15}$$

where

$$R_1=-\int_0^1 g(\xi)d\xi$$
$$\vdots$$
$$R_n=(-1)^n\int_0^1\int_0^{\xi_1}\cdots\int_0^{\xi_{n-1}}(1-\xi_1+\xi_n)g(\xi_1)(\xi_1-\xi_2)g(\xi_2)\cdots(\xi_{n-1}-\xi_n)g(\xi_n)d\xi_n\cdots d\xi_1. \tag{4.16}$$

As $\operatorname{Det} u(T)=1$, the only nontrivial invariant of $u(T)$ is (4.15)$\Rightarrow$the algebraic properties of the unimodular matrix $u(T)$ depend exclusively upon the function $R(\alpha)$ of $\alpha=aT^2$. We shall call it the *loop function*. (As $g(\xi)$ is bounded in $[0,1]$, the $R(\alpha)$ is an entire function of α.) A classification of the motions according to the value of $R(\alpha)$ follows. If $|R(\alpha)|<2$, the matrix $u(T)$ has two different complex eigenvalues, both of absolute value 1. Hence, the multiple repetitions of the evolution transformation induced by one period of $\gamma(t)$ yield a confined trajectory in the q,p space ("stability domain"). In addition, the real numbers α_{nl} for which $R(\alpha_{nl})=2\cos(2\pi l/n)$ $(n=3,4,\ldots,0<2l<n)$ represent such special values of $\alpha=aT^2$ for which $u(T)^n=1$, and hence, the operation induced by n oscillations of $\gamma(t)$ is an evolution loop. The values of α with $|R(\alpha)|=2$ correspond to a degenerate $u(T)$. If $|R(\alpha)|>2$, $u(T)$ acquires two real roots λ, $1/\lambda$ $(\lambda\neq0)$, and since one of them has the absolute value >1, the Heisenberg motion trajectory in q,p space is deconfined (in the classical limit it means the deconfinement of almost all classical phase space trajectories).

Suppose now, that $R_1=-\int_0^1 g(\xi)d\xi\neq0$. Then $R(0)=2$ implies that $R(\alpha)$ intersects $(-2,2)$ for $|\alpha|$ small enough with $\alpha R_1<0$. Hence, the potential $V(q,t)=\gamma(t)q^2/2=ag(t/T)q^2/2$, with small $|a|$, causes the confined motions and evolution loops when the sign(a) assures that the attractive impulses of $\gamma(t)$ in $[0,T]$ "prevail" over the repulsive ones. Less obvious cases of confinement, similar to the loops of Sec. III are obtained for $\gamma(t)$ with vanishing time average:

$$\int_0^T\gamma(t)dt=0\Leftrightarrow\int_0^1 g(\xi)d\xi=0. \tag{4.17}$$

Then $R_1=0$, and after putting $\phi(\xi)=\int_0^\xi g(\zeta)d\zeta$, R_2 reduces to

$$R_2=\left[\int_0^1\phi(\xi)d\xi\right]^2-\int_0^1\phi(\xi)^2\,d\xi<0. \tag{4.18}$$

This means that for sufficiently small $|\alpha|\neq0$, $R(\alpha)\cong2-|R_2|\alpha^2\Rightarrow|R(\alpha)|<2$. Thus we have the following lemma.

Lemma 3: If (4.17) holds, then for $|\alpha|\neq0$ small enough, the potential $V(q,t)=\gamma(t)q^2/2=ag(t/T)q^2/2$ generates the confined motions and evolution loops in $[0,nT]$ $(n=1,2,3,\ldots)$.

What is curious about this lemma is that, if (4.17) holds, then in any $[0,nT]$ the function $\gamma(t)$ provides as much attractive as repulsive impulses to the particle. One might think that they cannot alter the motion "in average." Yet, this is not the case. If $R(aT^2)$ takes one of the critical values, the free motion is effectively blocked. Instead of "dissolving," Schrödinger's wave is trapped into a kind of "periodic dance."

Displacement Lemma: A certain general property of the periodic loops is worth noticing. Suppose $\gamma(t)$ has a period τ, and the potential $V(q,t)=\gamma(t)q^2/2$ generates an evolution loop in $[0,\tau]$. Then take any $\sigma\in(0,\tau)$ and decompose the evolution operator $U(0,\tau)$ into $U(0,\tau)=U(0,\sigma)U(\sigma,\tau)$. The loop identity $U(0,\tau)\equiv1$ means $U(0,\sigma)U(\sigma,\tau)\equiv1$. Multiplying both sides by $U(0,\sigma)^{-1}$ from the left and by $U(0,\sigma)$ from the right, and using $U(0,\sigma)=U(\tau,\tau+\sigma)$ (caused by the periodicity of γ), one has

$$1\equiv U(\sigma,\tau)U(0,\sigma)=U(\sigma,\tau)U(\tau,\tau+\sigma)\equiv U(\sigma,\tau+\sigma).$$

Hence, we have the following lemma.

Lemma 4: If a periodic $\gamma(t)$ with a period τ generates an evolution loop in $[0,\tau]$, then $\gamma(t)$ generates also a loop in any other interval $[\sigma,\tau+\sigma]$. Schrödinger's particle injected into the pulsating field $\gamma(t)q^2/2$, in no matter which time moment, always returns to its initial state after the time τ.

Lemma 4': If a periodic function $\gamma(t)$ with a period τ produces an evolution loop in the time interval $[0,\tau]$, so does $\gamma(t+\sigma)$ $(\sigma\in\mathbb{R})$.

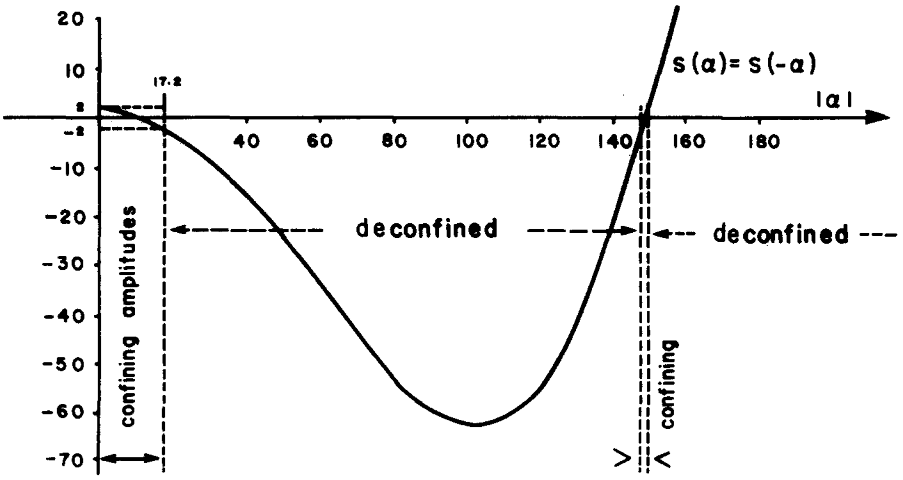


FIG. 3. The loop function $S(\alpha)$ obtained by an explicit numerical integration of (4.7).

C. Sinusoidally generated loops

Can the harmonic oscillations of $\gamma(t)$ produce the evolution loops? Evaluating (4.16) for $g(t) = \sin 2\pi t$, one obtains in place of $R(\alpha)$ a certain new special function $S(\alpha)$ of $\alpha = aT^2$, which yields a classification of "sinusoidally generated" motions (see Fig. 3):

$$S(\alpha) = 2 - \frac{1}{2}\frac{1}{(2\pi)^2}\alpha^2 + \frac{8\pi^2 - 75}{6144\pi^6}\alpha^4 + \cdots. \quad (4.19)$$

Consistent with Lemma 3, for $|\alpha|$ belonging to finite range of small amplitudes $0 < |\alpha| \lesssim 17.2 \Rightarrow |S(\alpha)| < 2$, the potential $a \sin(2\pi t/T)\, q^2/2$ creates the confined motions. The confinement interval $(-17.2, 17.2)$ is densely populated by the critical values α_{nl}: $S(\alpha_{nl}) = 2\cos(2\pi l/n)$ ($n = 3,4,\ldots, 0 < 2l < n$) and whenever α coincides with one of them, the evolution closes up after n periods of $\gamma(t)$. The particularly simple forms of cyclic motions with periods $3T$, $4T$, and $6T$ occur for $S(\alpha) = -1$, $S(\alpha) = 0$, $S(\alpha) = 1$ ($\alpha \cong \pm 14.8$, $\alpha \cong \pm 12$, $\alpha \cong \pm 8.4$). If $\alpha \neq \alpha_{nl}$ but $\alpha \in (-17.2, 17.2)$, the cyclic motion is desynchronized but remains trapped.

If $|\alpha|$ crosses 17.2 the picture changes. For $17.2 \lesssim |\alpha| \lesssim 148$, $S(\alpha) < -2$ and $\gamma(t)$ generates unlimited trajectories ("resonance band"). The confined and cyclic motions reappear in a narrow interval $148 \lesssim |\alpha| \lesssim 149.7$, where $S(\alpha)$ changes sign again. Due to farther increasingly sharp oscillations of $S(\alpha)$ around zero (computer materials; unpublished), the region $|\alpha| > 149.7$ is composed of a sequence of growing "resonance bands" encrusted with quickly shrinking confinement intervals. The question of whether this continues as $|\alpha| \to \infty$ is open.

A curious property of the "sinusoidal loops" follows from Lemma 4'. If $\gamma(t) = a \sin(2\pi t/T)$ generates a loop in $[0,\tau]$, $\tau = nT$, $n = 1,2,\ldots$, then the function $\gamma(t + \frac{1}{2}T) = -\gamma(t)$ must be loop generating, too. Hence, both potentials $\gamma(t)q^2/2$ and $-\gamma(t)q^2/2$ create the loops in the same time interval $[0,\tau]$. However, they can be viewed as the same electromagnetic field applied to two particles with identical masses and opposite charges. This exhibits a certain operational difference between the variable quadratic field $\gamma(t)q^2/2$ and the potential of the harmonic oscillator $(\omega^2/2)q^2$ [with the time-independent Hamiltonian $H_0 = \frac{1}{2}p^2 + (\omega^2/2)q^2$]. The harmonic oscillator also produces loops (the operator e^{-itH_0} returns to 1 for any $t = nT$, $T = 2\pi/\omega$). However, this holds only for Schrödinger's particle with a given charge. When the charge changes the sign, the loop effect disappears. Hence, the oscillating field $V(q,t) = \gamma(t)q^2/2$ [$\gamma(t) = a \sin \omega t$], with α in the confinement domain, is operationally superior over $\omega^2 q^2/2$, as it is able to trap both negative and positive charges.

Returning to the traditional units with $V(x,t) = ea \sin(2\pi t/T)x^2/2$, one has to put $\alpha = eaT/m$ (e, m being the charge and the mass of Schrödinger's particle), and henceforth, our main range of confined motions is characterized by a special relation between the charge/mass ratio, the period and the amplitude of $\gamma(t)$:

$$|(e/m)aT^2| \lesssim 17.2. \quad (4.20)$$

If the particle in question is Schrödinger's electron, $(e/m) \cong (4.77/9.03)\ 10^{18}\ \mathrm{cm}^{3/2}\,\mathrm{g}^{-1/2}\,\mathrm{sec}^{-1}$, and if γ oscillates with a frequency of short radio waves ($T \cong 10^{-7}$ sec), the lowest-range amplitudes a generating the loops correspond to the quadratic potentials $(a \sin(2\pi t/T))x^2/2$, whose maxima at the distance $x = 1$ meter from the center of the operation $x = 0$ are of the order of magnitude $\approx \alpha 10$ V $\lesssim 17.2 \times 10$ V. Similar evaluation can be done for the subsequent confinement intervals.

The phenomena that occur out of the confinement domain present a separate problem. It is possible that they correspond to a new type of resonance, which appears in the semiclassical theory, and represents the response of the particle to coherent fields instead of single quanta.

D. The loops for $V(q,t) = (a \sin(2\pi n/T)\, t)(q^2/2)$, $n = 1,2,3,\ldots$

Consider now the same interval $[0,T]$ but a new modulating function $\gamma_n(t) = a \sin(2n\pi t/T)$. The corresponding evolution operator $U(0,T)$ is $U(0,T/n)^n$, where $U(0,T/n)$ is the unitary operator generated by one oscillation period of $\gamma_n(t)$ (in the interval $[0,T/n]$). The same relation exists between the corresponding transformation matrices,

$$u(0,T) = u(0,T/n)^n, \quad (4.21)$$

and since they are 2×2 and unimodular, the trace of the nth power

$$S_n(\alpha) = \mathrm{Tr}[u(0,T)] = \mathrm{Tr}[u(0,T/n)^n]$$

can be easily expressed in terms of

$$\mathrm{Tr}[u(0,T/n)] = S[a(T/n)^2] = S(\alpha/n^2),$$

giving the following formulas for the "higher-order loop functions" $S_n(\alpha)$:

$$\begin{aligned}
&S_1(\alpha) = S(\alpha),\\
&S_2(\alpha) = S(\alpha/4)^2 - 2,\\
&S_3(\alpha) = S(\alpha/9)[S(\alpha/9)^2 - 3],\\
&S_4(\alpha) = [S(\alpha/16)^2 - 2]^2 - 2,\\
&\vdots\,.
\end{aligned} \tag{4.22}$$

The points α, where each of them intersects the values $-1,0,1,\ldots,2\cos(2\pi l/k),\ldots$, define the evolution loops of the oscillation moods $\gamma_n(t)$.

E. Simultaneous loop effects

In turn, take the modulating function $\gamma(t)$ composed of two oscillation moods,

$$\gamma_{nm}(t) = a\sin(2\pi nt/T) + b\sin(2\pi mt/T), \tag{4.23}$$

and consider the evolution operator in $[0,T]$ generated by the corresponding variable potential $\gamma_{mn}(t)q^2/2$. The trace of the associated evolution matrix now will be a function of two dimensionless variables, which is most convenient to choose as $\alpha = aT^2/n$, $\beta = bT^2/m$,

$$\mathrm{Tr}[u(0,T0)] = S_{nm}(\alpha,\beta). \tag{4.24}$$

The structure of this function for n odd and m even suggests some new manipulation possibilities. Indeed, consider the loop function $S(\alpha,\beta) = S_{12}(\alpha,\beta)$. A few first terms of (4.14) yield

$$\begin{aligned}
S(\alpha,\beta) = 2 &- (1/8\pi^2)(\alpha^2+\beta^2)\\
&+ (2^3\pi^2 - 3\cdot 5^2)3^{-1}2^{-11}\pi^{-6}\alpha^4\\
&+ (2^5\pi^2 - 3\cdot 5^2)3^{-1}2^{-13}\pi^{-6}\beta^4\\
&- (65 - 6\pi^2)3^{-2}2^{-8}\pi^{-6}\alpha^2\beta^2 + \cdots.
\end{aligned} \tag{4.25}$$

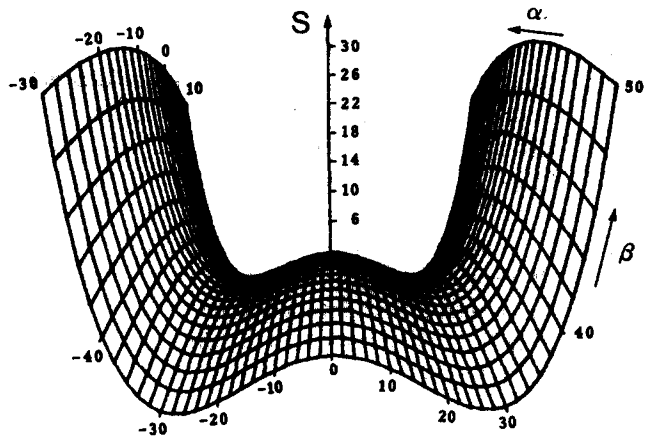


FIG. 4. The behavior of $S(\alpha,\beta)$ for low amplitudes.

The behavior of $S(\alpha,\beta)$ for low amplitudes, in $D=\{(\alpha,\beta): |\alpha|, |\beta|<148\}$, is of interest. For $\beta=0$, $S(\alpha,\beta)$ reduces to $S(\alpha,0)=S(\alpha)$, while for $\alpha=0$ it becomes $S(0,\beta)=S_2(2\beta)=S(\beta/2)^2-2$. Since $S(\alpha)<-2$ and $S_2(2\beta)>2$ for $|\alpha|\gtrsim 17.2$, $|\beta|>2.17.2$, hence, on every circle $\alpha^2+\beta^2=\rho^2$ $(2.17.2<\rho<148)$, $S(\alpha,\beta)$ has to accept all the values in $[-2,2]\Rightarrow$the curve S^κ defined by $S(\alpha,\beta)=\kappa(|\kappa|<2)$ intersect every circle $\alpha^2+\beta^2=\rho^2$ $(2.17.2<\rho<148)$. However, if $\kappa\to 2$, the equation $S(\alpha,\beta)=\kappa$ has an obvious solution in form of a closed, almost circular curve S_0^κ in the vicinity of (0,0) approximately given by $\alpha^2+\beta^2\cong(2-\kappa)8\pi^2$. Hence, S^κ must have another branch. In fact, this is confirmed by the computer picture of $S(\alpha,\beta)$ (Fig. 4) and the corresponding picture of the curves S^κ (Fig. 5). As one can see, for $-2<\kappa<2$, the S^κ is composed of three disjoint branches: S_0^κ (the one surrounding $\alpha=\beta=0$), and the S^κ_- and S^κ_+, whose arms escape from the computation region in the direction of β' as negative and positive, respectively. This decomposition implies the existence of oscillation patterns able to

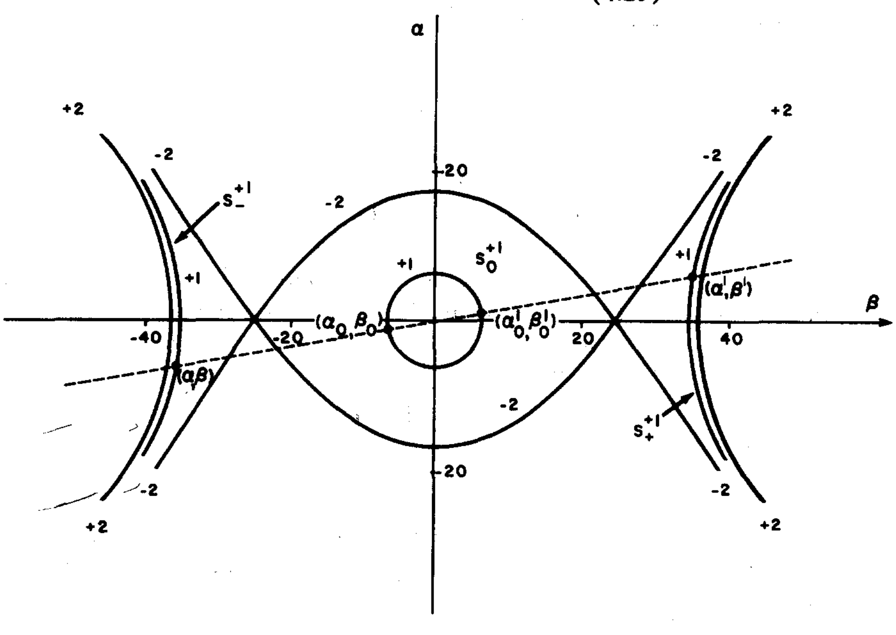


FIG. 5. The map of curves $S(\alpha,\beta)=$ const determined by a straightforward numerical integration of the matrix equation (4.7) with $\gamma(t)=\alpha\sin 2\pi t + 2\beta\sin 4\pi t$.

trap simultaneously four different Schrödinger's particles with different charge/mass ratios. Indeed, choose one of the values $\kappa = \kappa_{nl} = 2\cos(2\pi l/n) \in (-2,2)$. Then, take a point $(\alpha,\beta) \in S^{\kappa}_{+} \cup S^{\kappa}_{-}$ and draw the straight line through (α,β) and the coordinate center $(0,0)$. It will intersect S^{κ} in four points: (α,β), $(\alpha',\beta') \in S^{\kappa}_{+} \cup S^{\kappa}_{-}$, and (α_0,β_0), $(\alpha'_0,\beta'_0) \in S^{\kappa}_{0}$. They determine four possible values of the charge/mass ratio corresponding to four Schrödinger's particles simultaneously trapped in a synchronized loop motion. If one expresses the amplitudes in the traditional units: $a = \alpha(m/e)1/T^2$, $b = \beta(m/e)1/T^2$, the four possible charge/mass ratios of trapped particles are

$$e/m, -e/\lambda m, -e/\Lambda m, e/\Lambda' m, \tag{4.26}$$

where λ,Λ,Λ' are defined as three positive proportionality coefficients between the four two-vectors on the α,β plane:

$$(\alpha,\beta) = -\lambda(\alpha',\beta') = -\Lambda(\alpha'_0,\beta'_0) = \Lambda'(\alpha_0,\beta_0). \tag{4.27}$$

Now, if $\kappa \to 2$, S^{κ}_{0} shrinks to $(0,0)$, whereas the $S^{\kappa}_{\pm}$ remain at a finite distance from it; henceforth, Λ and Λ' become arbitrarily large. In particular, by resolving the equations,

$$S(\alpha,\beta) = S\left(-\frac{\alpha}{\Lambda}, -\frac{\beta}{\Lambda}\right) = 2\cos\frac{2\pi l}{n}. \tag{4.28}$$

With $\Lambda = 1838.3 \pm 1.0$, one can find the pairs of the oscillation amplitudes trapping both electrons and protons in a synchronized cyclic motion of a period $\tau = nT$. This kind of effect never could be achieved by static fields. Neither could it be achieved as a simple resonance phenomenon, triggered by the mere presence of some definite frequency. With all reservations about the applicability of evolution equations with c-number external potentials, this leads to an idea about the existence of more involved response phenomena, induced by softly changing, coherent fields and consisting in a combinatorial reaction to the simultaneous presence of many distinct oscillation moods.

Our techniques, until now, yielded continuous analogs of the loop formulas of Sec. III. A technique of achieving the free evolution regress $e^{i\tau p^2/2}$ $(\tau > 0)$ as a solution of (4.1) with continuous $\gamma(t)$ is also provided.[26] The use of the evolution loops to generate arbitrary unitary operations in $L^2(\mathbb{R})$ is now worth discussion.

V. PRECESSION

The evolution loops occupy a special place in the manipulation theory, since they produce *precession phenomena*. Suppose a one-parameter family of unitary operations $\{U_t: t \in [0,\tau]\}$ is an evolution loop: $U_\tau = U_0 = 1$. After altering slightly the external potential, the loop, in general, will not close: the gap $U_\tau U_0^{-1} = U_\tau$ will measure what we call the "precession of the operator orbit" (Fig. 6).

The so-defined precession is one of purest cases of the "interaction scheme." Indeed, if the free motion of a system is perturbed, the evolution operator splits: the "interaction part" is always accompanied by the free evolution part. Hence, the "little operations" of the interaction scheme cannot be truly superposed. In contrast to that, if an evolution loop is perturbed, a natural time interval to consider the

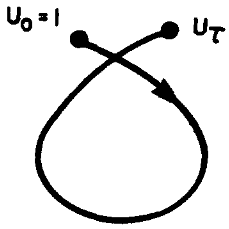
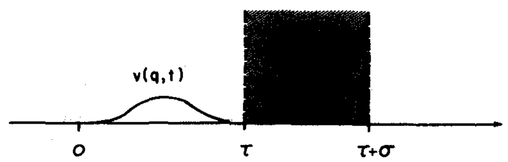


FIG. 6. Precession of the distorted loop.

result of the operation is the finite period of the loop motion. If one now splits the evolution into the zero part (cyclic) and the interaction part, then in any $[t_0, t_0+\tau]$ the zero part becomes the identity, and the evolution operator reduces to the pure interaction part. Hence, the "small operations" of the interaction picture can be effectively achieved. This suggests the following techniques of generating the unitary operations: (1) enclose the particle in an evolution loop, and (2) perturb the loop achieving the required operation by adding "precession effects."

As an example, consider the simplest loop composed of an interval of free evolution $[0,\tau]$ and subsequent interval $[\tau,\tau+\sigma]$ in which the external fields induce $e^{i\tau p^2/2}$. Now suppose the loop is perturbed in its free part $[0,\tau]$ by an additional external potential $V(q,t)$:



The cycle then is not closed, and the evolution operator within $[0,\tau+\sigma]$ is

$$U = \exp\left(i\tau\frac{p^2}{2}\right) T\left\{\exp\left(-i\int_0^\tau H(t)dt\right)\right\}. \tag{5.1}$$

As immediately seen, this operator is identical to the evolution operator that was thought typical for the interaction picture:

$$\begin{aligned}\frac{dU}{d\tau} &= -i[e^{i\tau p^2/2}V(q,\tau)e^{-i\tau p^2/2}]U \\ &= -iV(q+\tau p,\tau)U.\end{aligned} \tag{5.2}$$

Hence, the traditional evolution operators of the interaction frame become genuine dynamic operations describing the precession effects. Assume now that the field $V(q,t)$ is weak. In agreement with the continuous Baker–Campbell–Hausdorff formula,[14–16,24] the evolution operator (5.1) and (5.2) is approximately

$$\begin{aligned}U(\tau) &= \exp[-i\Omega(\tau)] \\ &\cong \exp\left(-i\int_0^\tau V(q+tp,t)dt\right).\end{aligned} \tag{5.3}$$

How general is this operator? What "tendencies to evolve" can be evoked by little perturbations of the evolution loops of Sec. III?

To check the contents of (5.3) it is convenient to express $\Omega(\tau)$ in terms of a fixed basis of independent, Hermitian forms of q and p. It turns out that the proper basis is suggested by the form of Eq. (5.2). Define $(A,B)=\frac{1}{2}(AB+BA)$. Then, consider the multiple anticommutators

$$\underbrace{(p,(p,\dots(p}_{n},\phi(q)\dots),\quad n=0,1,\dots. \tag{5.4}$$

When ϕ runs through a basis of polynomials in q, and $n=0,1,2,\dots$, the expressions (5.4) become a basis of independent, Hermitian forms of p and q. Note now

$$[i\,(p^2/2)\,,\phi(q)]=\tfrac{1}{2}\,(p\phi'(q)+\phi'(q)p)=(p,\phi'(q)), \tag{5.5}$$

$$\vdots$$

$$\underbrace{[i\,(p^2/2)\,,[i\,(p^2/2)\,,\dots[i\,(p^2/2)}_{n}\,,\phi(q)]\dots]]$$

$$=(p,(p,\dots(p,\phi^{(n)}(q)\dots). \tag{5.6}$$

Hence

$$\begin{aligned}V(q+tp,t)&=e^{itp^2/2}V(q,t)e^{-itp^2/2}\\&=\sum_{n=0}^{\infty}\frac{t^n}{n!}\left[i\frac{p^2}{2},\left[i\frac{p^2}{2},\dots\left[i\frac{p^2}{2},V(q,t)\right]\dots\right]\right.\\&=\sum_{n=0}^{\infty}\frac{t^n}{n!}\left(p,\left(p,\dots\left(p,\frac{\partial^n}{\partial q^n}V(q,t)\right)\dots\right)\right.. \end{aligned}\tag{5.7}$$

and

$$\begin{aligned}\Omega(\tau)&\cong\int_0^\tau V(q+tp,t)dt\\&=\sum_{n=0}^{\infty}\frac{1}{n!}\left(p,\left(p,\dots\left(p,\frac{\partial^n}{\partial q^n}\int_0^\tau t^nV(q,t)dt\right)\dots\right)\right.,\end{aligned}\tag{5.8}$$

which yields the desired decomposition of the linear part of $\Omega(\tau)$ into independent anticommutators $(p,(p,\dots(p,\phi(q))\dots)$. If, in particular,

$$V(q,t)=\gamma(t)V(q), \tag{5.9}$$

then

$$\begin{aligned}\Omega(\tau)&\cong\sum_{m=0}^{\infty}\left[\frac{1}{m!}\int_0^\tau t^m\gamma(t)dt\right]\\&\quad\times\underbrace{\left(p,\left(p,\dots\left(p\right.\right.\right.}_{m},\left.\left.\left.\frac{d^mV}{dq^m}\right)\dots\right)\right)\\&=\sum_{m=0}^{\infty}\gamma_m\left(p,\left(p,\dots\left(p,\frac{d^mV}{dq^m}\right)\dots\right)\right),\end{aligned}\tag{5.10}$$

where the coefficients γ_m are simply the mathematical moments of $\gamma(t)$ in $[0,\tau]$. If, moreover, $V(q)$ is a polynomial of order N, the series (6.10) ends up after N terms, and since any first N moments of $\gamma(t)$ in $[0,\tau]$ are arbitrary, the exponent of the unitary operation (5.3) can be manipulated quite easily by $\gamma(t)$. Taking $\gamma(t)$ such that

$$\gamma_m=\frac{1}{m!}\int_0^\tau t^m\gamma(t)dt=\delta_{m,m_0}\quad(m=0,\dots,N), \tag{5.11}$$

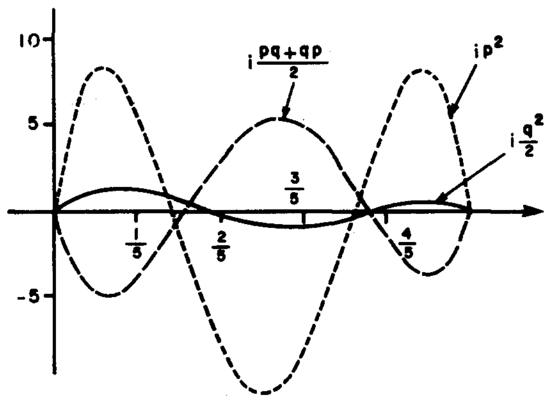


FIG. 7. Three effective exponents of the evolution induced by $\gamma(t)q^2/2$ as $\gamma(t)$ varies according to each plotted oscillation mood (approximation linear in γ).

one can reduce (6.3) to any one of the operators

$$e^{-iV(q)},e^{-i(p,V'(q))},\dots,e^{-i(p,(p,\dots(p,V^{(N)}(q))\dots)}. \tag{5.12}$$

[Figure 7 represents a system of three functions, yielding this kind of reduction for $V(q)=\frac{1}{2}q^2$.]

Since the $V^{(n)}(q)$ are arbitrary, (5.4) is a basis, and the "little precession exponents" of (5.12) can be added by the mechanism of Trotter formulas,[25] one arrives at the following lemma.

Lemma 5: The precession of the distorted loop (5.1) permits us to approach any operation $e^{-if(q,p)}$ with a Hermitian $f(q,p)$ as a formal solution of the evolution problem. (Note that the sense of this lemma is purely algebraic. The problems of the essential self-adjointness of the exponents and the strong convergence of the corresponding evolution operators remain open.)[20]

Some operational questions follow.

The present-day techniques of intervening into the structure of atomic and molecular systems relies mostly on simple resonance effects. However, Lemma 5 suggests some different possibilities.

(1) Given a Schrödinger particle in an oscillator field ($H_0=\frac{1}{2}p^2+(\omega^2/2)q^2$) can one design an external potential $V(q,t)$ transforming some eigenstates of H_0 only, without affecting the others? (Note that this could not be done just by bombarding the oscillator with an incoherent photon beam.)

(2) Given two different bound systems with the same distance between two basic energy levels, can one design an external potential $V(q,t)$ that would deconfine one of them without affecting the other?

If the answers were positive, within the applicability limits of quantum mechanics with the c-number $V(q,t)$, it might mean that the coherent fields, interacting with the evolution loops, are naturally suited to become the "selective pincers" of the particle technologies.

VI. LOOPS OF SOURCE-FREE FIELDS

The evolution problem in one space dimension is still far from operational. The manipulation of Schrödinger's particle by an arbitrary $V(q,t)$, in general, means the application of an electromagnetic four-potential A^α with nonvanishing sources. A more realistic scheme would be to consider the wave packet manipulated by free electromagnetic fields arriving from a distance. In the nonrelativistic approximation (consistent with Schrödinger's quantum mechanics) this means the application of the *harmonic* potentials in $\mathbb{R}^n$, $n > 1$:

$$\Delta V(\bar{x},t) = 0. \tag{6.1}$$

A question arises: can the manipulation prescriptions of Secs. III–V be repeated by using only source-free fields?

A. Loops of harmonic potentials in $\mathbb{R}^2$

Consider the loops of the quadratic potentials $\gamma(t)q^2/2$ in $L^2(\mathbb{R})$ for which the change of sign, $\gamma(t) \to -\gamma(t)$, does not affect the loop property (i.e., both $\gamma(t)q^2/2$ and $-\gamma(t)q^2/2$ produce the identity operation within the same time interval $[0,T]$). Then the following potential, which is harmonic in $\mathbb{R}^2$, must generate an evolution loop in $L^2(\mathbb{R}^2)$:

$$V(x,y,t) = \gamma(t)(x^2/2 - y^2/2). \tag{6.2}$$

Note that the required sign invariance is shared by a sequence of evolution loops (3.15)–(3.18) of Sec. III as well as by the sinusoidal loops (4.19) of Sec. IV. Henceforth, each of them defines a harmonic loop of form (6.2) in $\mathbb{R}^2$.

B. Harmonic loops in $\mathbb{R}^3$

The loops of Sec. III permit us, as well, to build up patterns of loop-creating harmonic pulses in $\mathbb{R}^3$. A simple case is obtained by considering a process whose repeated sequence consists of three consecutive pulses of the harmonic potentials $(x^2 - y^2)/2$, $(y^2 - z^2)/2$, $(z^2 - x^2)/2$ separated by three rest intervals of the length τ. The resulting evolution operation $U = U(0,3\tau)$ is

$$U = e^{-i\tau\bar{p}^2/2} e^{-ia(z^2 - x^2)/2} e^{-i\tau\bar{p}^2/2} e^{-ia(y^2 - z^2)/2} \times e^{-i\tau\bar{p}^2/2} e^{-ia(x^2 - y^2)/2}, \tag{6.3}$$

and it splits into a product of three commuting "partial evolution operations" describing the propagation in the x, y, and z directions:

$$\begin{aligned} U &= W(x,p_x)\Lambda(y,p_y)S(z,p_z) \\ &= \left[e^{-i\tau p_x^2/2} e^{aix^2/2} e^{-2i\tau p_x^2/2} e^{-iax^2/2}\right] \\ &\quad \times \left[e^{-2i\tau p_y^2/2} e^{-iay^2/2} e^{-i\tau p_y^2/2} e^{iay^2/2}\right] \\ &\quad \times \left[e^{-i\tau p_z^2/2} e^{-iaz^2/2} e^{-i\tau p_z^2/2} e^{iaz^2/2} e^{-i\tau p_z^2/2}\right]. \end{aligned} \tag{6.4}$$

Now, forget about the concrete names of the variables x, y, and z and consider an abstract sequence:

$$W(q,p) = e^{-i\tau p^2/2} e^{iaq^2/2} e^{-2i\tau p^2/2} e^{-iaq^2/2}, \tag{6.5}$$

where q,p can mean any of the pairs x_j,p_j ($j = 1,2,3$). When $a = 1/\tau$, $W(q,p)$ becomes a part of the loop of Fig. 2, and so

$$W^4 = [e^{-i\tau p^2/2} e^{i(1/\tau)q^2/2} e^{-2i\tau p^2/2} e^{-i(1/\tau)q^2/2}]^4 \equiv 1. \tag{6.6}$$

Now, if A, B are any operators and A is invertible, then

$$(AB)^n \equiv 1 \Rightarrow B(AB)^{n-1}A \equiv 1 \Rightarrow (BA)^n \equiv 1. \tag{6.7}$$

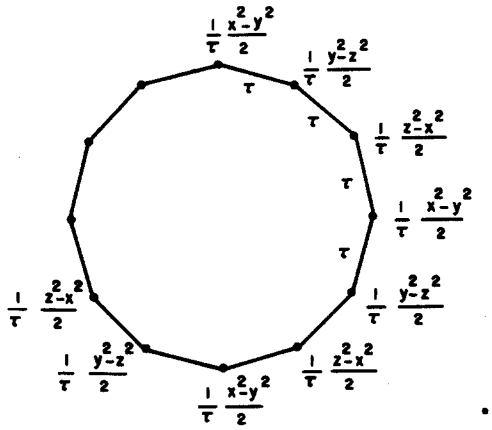
Hence, the components of (6.6) inside of the square bracket may be cyclically permuted yielding

$$W(q,p)^4 \equiv 1 \Rightarrow \Lambda(q,p)^4 \equiv S(q,p)^4 \equiv 1. \tag{6.8}$$

Thus, for U defined by (6.5) and $a = 1/\tau$,

$$U^4 \equiv 1, \tag{6.9}$$

which defines one of the simplest harmonic loops in $L^2(\mathbb{R}^3)$:



The harmonic equivalents in $L^2(\mathbb{R}^3)$ of other loop formulas (3.17) with $t = 2\tau$ are similarly obtained.

C. The simplest assymetric sign-invariant loop

The three-dimensional harmonic analog of the loop (3.17), used to generate the free evolution inversion $e^{i\tau\bar{p}^2/2}$, share the lack of "time economy" characteristic for the regular loops of Sec. III. Is there a harmonic but "time-economic" alternative to these prescriptions? To find an answer, it is convenient to go back to the evolution problem in $L^2(\mathbb{R})$ (one space dimension) and to ask the following question: can one have a sign-invariant alternative of the inversion prescription (3.24)? Differently than in Sec. III, the question is numerically nontrivial. Due to the number of demands to be satisfied, the techniques of three $q^2/2$-pulses is not sufficient. What one needs are at least six pulses of $q^2/2$ potential intertwined with six distinct intervals of free evolution, most generally leading to the following unitary operation in $L^2(\mathbb{R})$:

$$U = \prod_{j=6,5,\ldots,1} (e^{-ia_j q^2/2} e^{-i\alpha_j p^2/2}). \tag{6.10}$$

This corresponds to the transformation matrix

$$u\begin{pmatrix} a_1,\ldots,a_6 \\ \alpha_1,\ldots,\alpha_6 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ a_1 & 1 \end{pmatrix}\begin{pmatrix} 1 & -\alpha_1 \\ 0 & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 \\ a_6 & 1 \end{pmatrix}\begin{pmatrix} 1 & -\alpha_6 \\ 0 & 1 \end{pmatrix}. \tag{6.11}$$

The demands that the product of the 12 unitary operations should give a loop, and that the loop property should not be affected by the simultaneous change of the sign of all a_j's ($j = 1,\ldots,6$) can be written as

$$u\begin{pmatrix} a_1,\ldots,a_6 \\ \alpha_1,\ldots,\alpha_6 \end{pmatrix} = u\begin{pmatrix} -a_1,\ldots,-a_6 \\ \alpha_1,\ldots,\alpha_6 \end{pmatrix} = 1, \tag{6.12}$$

and they provide six independent algebraic equations for the six real pulse amplitudes $a_1,\ldots,a_6$. For arbitrary $\alpha_1,\ldots,\alpha_6$ the solution requires computer techniques. However, for some special collections of the "rest intervals" $\alpha_1,\ldots,\alpha_6$, explicit algebraic solutions exist. In what follows, it will be convenient to put $\alpha_1 = \alpha, \alpha_2 = \beta, \alpha = \gamma, \alpha_4 = \alpha', \alpha_5 = \beta', \alpha_6 = \gamma'$, $a_1 = a, a_2 = b, a_3 = c, a_4 = a', a_5 = b', a_6 = c'$ (see Fig. 8).

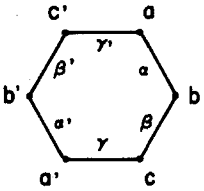


FIG. 8. The general six-pulse loop.

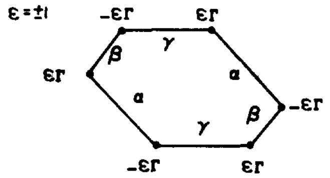


FIG. 9. A generalization of the sign-invariant loop of Sec. III. Here $\epsilon = \pm 1$ and $\Gamma = [(\alpha+\beta+\gamma)/\alpha\beta\gamma]^{1/2}$.

Then, the soluble cases are as follows.

Case 1: If $\alpha' = \alpha$, $\beta_1 = \beta$, $\gamma' = \gamma$, then consistent with the required sign invariance and cyclic symmetry (6.8), it is plausible to assume $a' = -a$, $b' = -b$, $c' = -c$. After substituting to (6.12) it yields

$$\left.\begin{aligned} a &= \epsilon\Gamma \\ b &= -\epsilon\Gamma \\ c &= \epsilon\Gamma \end{aligned}\right\} \Rightarrow \begin{aligned} a' &= -\epsilon\Gamma, \\ b' &= \epsilon\Gamma, \\ c' &= -\epsilon\Gamma, \end{aligned}$$
$$\Gamma = \sqrt{(\alpha+\beta+\gamma)/\alpha\beta\gamma}, \quad \epsilon = \pm 1, \tag{6.13}$$

henceforth, providing a generalization of the sign-invariant loop of Sec. III (see Fig. 9).

This loop, however, does not allow us to achieve the operation $e^{i\alpha p^2/2}$ within a time shorter than $\alpha + 2\beta + 2\gamma > \alpha$, thus failing to provide the required "time economy."

Case 2: Suppose $\alpha' = \beta$, $\beta' = \alpha$. Then the sides of the loop diagram together with the cyclic and sign-invariance properties suggest $c' = -a$, $b' = -b$, $a' = -c$, which, after feeding up to (6.12) and denoting $\gamma = \tau$, $\gamma' = t$, yields

$$\begin{aligned} a &= \frac{\epsilon\sqrt{\Sigma}}{\sqrt{\alpha t(\alpha+\beta)}} = -c', \quad \epsilon = \pm 1, \\ b &= -\frac{\epsilon\sqrt{\Sigma}\sqrt{\alpha+\beta}}{\sqrt{\alpha}\sqrt{\beta}\,(\sqrt{\alpha\tau}+\sqrt{\beta\tau})} = -b', \\ c &= \frac{\epsilon\sqrt{\Sigma}}{\sqrt{\alpha}\sqrt{\beta}\,\sqrt{\alpha+\beta}} = -a', \quad \Sigma = 2\alpha + 2\beta + t + \tau. \end{aligned} \tag{6.14}$$

This is the required sign-invariant alternative of the asymmetric loop of formula (3.25). In particular, taking $\alpha = \beta = \gamma = \alpha' = \beta' = \tau$ and $\gamma' = t$ one obtains the simplest possible sign-invariant and asymmetric loop represented in Fig. 10.

The corresponding loop identity can be equivalently written as

$$\begin{aligned} \exp\left(it\frac{p^2}{2}\right) \equiv{}& \exp\left(\frac{i\Lambda}{\sqrt{t}}\frac{q^2}{2}\right)\exp\left(-i\tau\frac{p^2}{2}\right)\exp\left(-\frac{2i\Lambda}{\sqrt{t}+\sqrt{\tau}}\frac{q^2}{2}\right)\exp\left(-i\tau\frac{p^2}{2}\right)\exp\left(\frac{i\Lambda}{\sqrt{\tau}}\frac{q^2}{2}\right) \\ &\times\exp\left(-i\tau\frac{p^2}{2}\right)\exp\left(-\frac{i\Lambda}{\sqrt{\tau}}\frac{q^2}{2}\right)\exp\left(-i\tau\frac{p^2}{2}\right)\exp\left(\frac{2i\Lambda}{\sqrt{t}+\sqrt{\tau}}\frac{q^2}{2}\right)\exp\left(-i\tau\frac{p^2}{2}\right)\exp\left(-\frac{i\Lambda}{\sqrt{t}}\frac{q^2}{2}\right). \end{aligned} \tag{6.15}$$

Due to the number of demands involved, this is the simplest existing assymetric, sign-invariant "inversion formula" in $L^2(\mathbf{R})$ based on electric potentials. By taking the numbers at the vertices of Fig. 10 to modulate the δ-singularities of $\gamma(t)$ in (6.2), one simultaneously obtains the simplest prescription for achieving the free evolution inversion in $L^2(\mathbf{R}^2)$ at the cost of δ-like harmonic pulses in an arbitrary short time.

D. Harmonic time inversion in $\mathbf{R}^3$

In $\mathbf{R}^3$ three obvious analogs of the inversion formula (6.15) are obtained by associating the pulse function $\gamma(t)$ of the loop of Fig. 10 with the three harmonic potentials $x^2/2 - y^2/2$, $y^2/2 - z^2/2$, and $z^2/2 - x^2/2$. In each case, the free evolution inversion in $L^2(\mathbf{R}^3)$ is incomplete, leaving one of the propagation directions free. The corresponding evolution operators are

$$\gamma(t)(x^2-y^2)/2 \to \exp\left(it\frac{p_x^2+p_y^2}{2}\right)\exp\left(-5i\tau\frac{p_z^2}{2}\right),$$
$$\gamma(t)(y^2-z^2)/2 \to \exp\left(it\frac{p_y^2+p_z^2}{2}\right)\exp\left(-5i\tau\frac{p_x^2}{2}\right),$$
$$\gamma(t)(z^2-x^2)/2 \to \exp\left(it\frac{p_z^2+p_x^2}{2}\right)\exp\left(-5i\tau\frac{p_y^2}{2}\right).$$

Hence, after executing them in sequence in the time interval $[0,15\tau]$, the complete evolution operator in $L^2(\mathbf{R}^3)$ becomes

$$U(0,15\tau) = e^{i(2t-5\tau)\bar{p}^2/2}, \tag{6.16}$$

and as t and τ are independent, taking $t = \text{const}$, $\tau \to 0$, one arrives at the following lemma.

Lemma 6: An arbitrarily deep regress of the free evolution in $L^2(\mathbf{R}^3)$ can be achieved by harmonic pulses acting in

$\Lambda = \pm \frac{1}{\tau}\sqrt{\frac{\Sigma}{2}}$

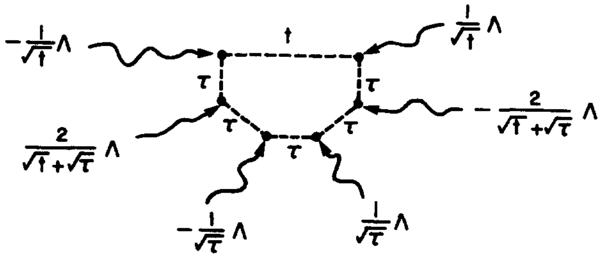


FIG. 10. The loop of formula (6.15).

an arbitrarily short time interval.

E. Waves trapped in continuous electric source-free field

The harmonic loops described up to now are induced by potential shocks. To achieve them, the experimenter would have to possess a method of sharpening indefinitely the field pulses. Even if technically possible, at some moment this would become counterproductive, as for too strong and too quickly varying fields, the validity of the quantum mechanical level of the theory ends up, and competitive dynamic phenomena are activated (the radiative corrections and then, the paircreation predicted by quantum field theory). Hence, the question arises: can the evolution loops be generated by *nonsingular* harmonic fields in $\mathbb{R}^3$?

Suppose $\gamma(t)q^2/2$ is one of quadratic potentials generating a loop in one-space dimension, and let $\gamma(t)$ be of the particular form

$$\gamma(t)=\phi(t)-\phi(t+\tfrac{1}{3}T), \tag{6.17}$$

where $\phi(t)$ is a periodic function with the period T. [This simply means that $\gamma(t)$ does not contain Fourier components with $n=3k$, $k=0,1,\ldots$.] Then define

$$V(\bar{x},t)=\gamma(t)\frac{x^2}{2}+\gamma\Big(t+\frac{1}{3}T\Big)\frac{y^2}{2}+\gamma\Big(t+\frac{2}{3}T\Big)\frac{z^2}{2}. \tag{6.18}$$

Due to (6.17), $V(\bar{x},t)$ is harmonic. Moreover, the evolution operator associated with it splits into three commuting parts, and, due to Lemma 4 applied for $\sigma=\frac{1}{3}T$ and $\sigma=\frac{2}{3}T$, they generate the simultaneous evolution loops for the three tensor product components of $\psi(\bar{x},t)$. Hence, (6.18) generates a loop in $L^2(\mathbb{R}^3)=L^2(\mathbb{R})\otimes L^2(\mathbb{R})\otimes L^2(\mathbb{R})$. Note that any $\gamma(t)=a\sin(2\pi nt/T)$ $(n\neq 3k)$ is of the required form (6.17); henceforth, the following harmonic potential, curiously reminding the three phase electric current

$$V(\bar{x},t)=a\sin\frac{2\pi nt}{T}\frac{x^2}{2}+a\sin\frac{2n\pi(t+\frac{1}{3}T)}{T}\frac{y^2}{2}+a\sin\frac{2n\pi(t+\frac{2}{3}T)}{T}\frac{z^2}{2}, \tag{6.19}$$

generates the loops in $L^2(\mathbb{R}^3)$ exactly for the values of the dimensionless amplitude $\alpha=(e/m)aT^2$ predicted for the loop motion in one space dimension. As in Sec. IV the resulting loops have the property of trapping simultaneously Schrödinger's particles with opposite charges into a synchronized periodic motion.

Now, the construction (6.18) can be repeated for the two-mood amplitude $\gamma(t)=a\sin(2\pi t/T)+b\sin(4\pi t/T)$ [which too, fulfills (6.17)]. The corresponding harmonic field (6.18) then permits us to transplant to $\mathbb{R}^3$ the effects discussed in Sec. IV. Thus, if the pair of the dimensionless amplitudes $\alpha=(e/m)aT^2, \beta=2b(e/m)T^2$ fulfills the conditions (4.28), the harmonic potential (6.18) grants the simultaneous loop motions for four types of particles with four different values of charge/mass. This, on a classical level, means the simultaneously closed phase trajectories in the (4×6)-dimensional phase space of four classical mass points. By taking the solutions of (4.28) with $\Lambda\cong 1838.3$, one obtains a harmonic potential able to confine in a synchronized cyclic motion the particles with the charge/mass ratios of the electron and of the proton. The problem of engineering "simultaneous traps" for several kinds of particles is open.

The question of how to create the electric potentials with the required quadratic dependence on space coordinates is open. Since they are harmonic, they are in principle interpretable as a result of an outside influence. However, in $\mathbb{R}^3$, much simpler manipulations are due to the existence of magnetic fields.

F. Nonsingular magnetic loop

Consider a charged Schrödinger particle moving in a homogeneous, time-dependent magnetic field $\bar{H}=2\gamma(t)\bar{n}$, where $\bar{n}$ is a unit vector and $\gamma(t)$ a function. The vector potential may be chosen as

$$\bar{A}=-\gamma(t)\bar{r}\times\bar{n},\quad \bar{r}=\begin{pmatrix}x\\y\\z\end{pmatrix}, \tag{6.20}$$

and the time-dependent Hamiltonian in the dimensionless coordinates becomes

$$\begin{aligned}H(t)&=\tfrac{1}{2}[\bar{p}+\gamma(t)\bar{r}\times\bar{n}]^2\\&=\tfrac{1}{2}\bar{p}^2-\gamma(t)\bar{n}\bar{M}+\gamma(t)^2\bar{r}_\perp^2/2,\end{aligned} \tag{6.21}$$

where $\bar{r}_\perp:=\bar{r}-(\bar{n}\bar{r})\bar{n}$, $\bar{M}=\bar{r}\times\bar{p}$. [We take a nonrelativistic approximation for A^α, consistent with Schrödinger's quantum mechanics. The particular form of (6.20) corresponds to a homogeneous magnetic pulse associated with an axially symmetric electric field.] To illustrate the creation of magnetic loops, choose $\gamma(t)$ to be a rectangular step function (see Fig. 11)

$$\gamma(t)=\begin{cases}+a, & \text{if } t\in[0,T/2],\\-a, & \text{if } t\in[T/2,T],\\0, & \text{otherwise.}\end{cases} \tag{6.22}$$

If $\bar{n}=$ const, the corresponding Hamiltonian (6.21) is a sum of two commuting families $H_1(t)=\frac{1}{2}\bar{p}^2+\gamma(t)^2\frac{1}{2}\bar{r}_\perp^2$ and $H_2(t)=-\gamma(t)\bar{n}\bar{M}$ and, since $\int_0^T\gamma(t)dt=0$, the evolution operator within the time interval $[0,T]$ is only due to $H_1(t)$. Taking $\bar{n}=[0,0,1]$, one has

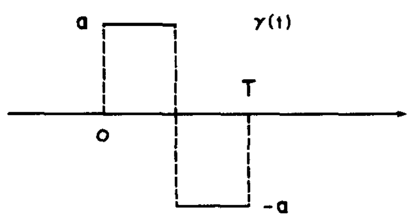


FIG. 11. The modulating function (6.22).

$$U_{\bar n} = \exp\left(-i\int_0^T H_1(t)dt \right)$$
$$= \exp\left(-iT\left[\frac{\bar p^2}{2} + a^2\frac{\bar r_\perp^2}{2}\right]\right)$$
$$= \exp\left(-iT\left[\frac{p_x^2+p_y^2}{2} + a^2\frac{x^2+y^2}{2}\right]\right)$$
$$\times \exp\left(-iT\frac{p_z^2}{2}\right). \qquad (6.23)$$

In turn, taking three consecutive step functions, $\gamma(t)$, $\gamma(t-T)$, and $\gamma(t-2T)$, and associating them with the sequence of three mutually orthogonal pulses of the magnetic field in the directions $\bar n = [0,0,1]$, $\bar m = [0,1,0]$, and $\bar s = [1,0,0]$,

$$\bar H(t) = \gamma(t)\bar n + \gamma(t-T)\bar m + \gamma(t-2T)\bar s, \qquad (6.24)$$

one obtains the following evolution operator in $[0,\tau]$, $\tau = 3T$:

$$U(\tau) = U_{\bar s} U_{\bar m} U_{\bar n}$$
$$= \exp\left(-iT\frac{p_x^2}{2}\right)\exp\left[-2iT\left(\frac{p_x^2}{2} + a^2\frac{x^2}{2}\right)\right]$$
$$\times \exp - \left[-iT\left(\frac{p_y^2}{2} + a^2\frac{y^2}{2}\right)\right]\exp\left(-iT\frac{p_y^2}{2}\right)$$
$$\times\left[-iT\left(\frac{p_y^2}{2} + a^2\frac{y^2}{2}\right)\right]$$
$$\times \exp\left[-2iT\left(\frac{p_z^2}{2} + a^2\frac{z^2}{2}\right)\right]\exp\left(-iT\frac{p_z^2}{2}\right)$$
$$= W(x,p_x)P(y,p_y)S(z,p_z). \qquad (6.25)$$

As in (6.4), forget about the names of the variables, and consider the algebraic elements of the process:

$$A = \exp\left[-iT\left(\frac{p^2}{2} + a^2\frac{q^2}{2}\right)\right], \quad B = \exp\left(-iT\frac{p^2}{2}\right). \qquad (6.26)$$

The typical sequences occurring in (6.25) are

$$W = BA^2, \quad P = ABA, \quad S = A^2B, \qquad (6.27)$$

and, due to the possibility of the cyclic permutations,

$$(W^n \equiv 1) \Rightarrow (P^n \equiv 1) \Rightarrow (S^n \equiv 1). \qquad (6.28)$$

Henceforth, to find the loop cases of (6.25) in $L^2(\mathbf{R}^3)$ it is enough to examine the structure of W. The corresponding 2×2 matrix is

$$\omega = \begin{pmatrix} 1 & -T \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \cos aT, & -\sin aT/a \\ a\sin aT, & \cos aT \end{pmatrix}, \qquad (6.29)$$

and

$$Tr\,\omega = r(aT) = 2\cos aT - aT\sin aT. \qquad (6.30)$$

Henceforth, the values of $\alpha = aT$ that fulfill the equations

$$r(\alpha) = 2\cos\alpha - \alpha\sin\alpha = 2\cos(2\pi l/n)$$
$$(n \geqslant 3, \quad 0 < 2l < n) \qquad (6.31)$$

define the cases for which $\omega^n = 1 \Rightarrow W^n \equiv P^n \equiv S^n \equiv 1 \Rightarrow U(\tau)^n \equiv 1$, and the magnetic field (6.24) produces an nth-order evolution loop in $L^2(\mathbf{R}^3)$. After plotting the even function $r(\alpha)$, one sees that for every $\kappa_{nl} = 2\cos(2\pi l/n) \in (-2,2)$, Eq. (6.31) is fulfilled for an infinite sequence of points on the α axis. They define special sequences of magnetic pulses, coming from three mutually perpendicular directions, which can trap Schrödinger's particle together with its charge conjugate into a perpetual cyclic motion. Note that a similar confinement never could be achieved in $\mathbf{R}^3$ by using static magnetic fields. [The magnetic confinement due to sinusoidal $\gamma(t)$ is also possible (to appear).]

G. The validity domain

Above, we have treated the quantum mechanics with time-dependent external potentials "as it is," which seems a legitimate task. However, the question of the applicability domain arises. As is clear, the discussed cyclic motions are only the simplest quantum mechanical models. In reality, the particle trapped into a loop must radiate, and, therefore, its motion must differ from the quantum mechanical trajectory. A question is how quickly will this happen? To obtain a rough estimate, assume the particle is classical charge e trapped in a cyclic motion of period T and amplitude l, and compare it with an oscillating dipole: $\bar d(t) = el\bar n\lambda(t/T)$, $\bar n$ being a unit vector and $\lambda(\xi)$ a periodic function of period 1. The dipole would radiate the power

$$p(t) \cong \frac{1}{c^3}|\ddot{\bar d}|^2 \cong \frac{1}{c^3}\frac{e^2l^2}{T^4}\left|\ddot\lambda\left(\frac{t}{T}\right)\right|^2,$$

causing the energy loss per period

$$E(T) \cong \frac{1}{c}\,3\,\frac{el}{T^4}\int_0^T \left|\ddot\lambda\left(\frac{t}{T}\right)\right|^2 dt$$
$$\cong \frac{1}{(cT)^3}e^2l^2\int_0^1 |\ddot\lambda(\xi)|^2 d\xi.$$

Assuming that

$$\int_0^1 |\ddot\lambda(\xi)|^2 d\xi \cong 1, \quad E(T) \cong \frac{e^2l^2}{(cT)^3},$$

and if the length of the particle trajectory within one period is $\cong l$, the energy loss per element of trajectory (interpreted as the radiative self-force) is $F_{\rm rad} \cong le^2/(cT)^3$. On the other hand, the field forces sustaining the loop are of the order of magnitude $F_{\rm field} \cong eal$, and due to the loop condition $(e/m)aT^2 = \alpha$, $F_{\rm field} \cong \alpha ml/T^2$,

$$\frac{F_{\rm rad}}{F_{\rm field}} \cong \frac{1}{\alpha}\frac{e^2}{cT}\frac{1}{mc^2} \cong 10^{-7}\frac{\rm cm}{(cT)} \cong 10^{-14}\frac{\rm sec}{T}. \qquad (6.32)$$

Thus, for the loop frequencies comparable to those of the short radio waves ($cT \cong 10^3$cm), $F_{\rm rad}/F_{\rm field} \cong 10^{-10}$, the radiative corrections can be neglected, whereas for the frequencies $\cong 10^{12}$/sec, they would become significant.

An intriguing question is what happens to the loop solutions due to the radiation emission. The traditional quantum mechanical oscillator, with the constant potential, after radiating quanta, settles on the lowest possible energy level (ground state). For the time-dependent loop potentials, the particle cannot fall to the ground state since there is none. The formation of some new, quasicyclic patterns of motion is henceforth probable.

VII. PRECESSION OF CYCLIC MOTIONS DUE TO SOURCE-FREE EXTERNAL FIELDS

Following Sec. V we shall examine the harmonic loops perturbed by harmonic fields. Similarly as in Sec. V consider the simplest loop composed of an interval $[0,\tau]$ of the free evolution $e^{-i\tau\bar{p}^2/2}$ and then an interval $[\tau,T]$ in which the harmonic fields induce $e^{i\tau\bar{p}^2/2}$. Assume now that the process is deformed by an additional potential $V(\bar{x},t)$ inserted into the free part $[0,\tau]$. As observed in Sec. V the evolution operator in $[0,T]$ coincides with that of the traditional interaction picture. If $V(\bar{x},t)$ is small [put $V(\bar{x},t)\to\epsilon V(\bar{x},t)$], the approximate expression is

$$U(0,T) = \exp(-i\Omega)$$
$$\cong \exp\Big(-i\epsilon\int_0^\tau V(\bar{x}+t\bar{p},t)\,dt\Big) = \exp\Big(-i\epsilon\underset{1}{\Omega}\Big). \tag{7.1}$$

Now, however, $V(\bar{x},t)$ is harmonic. Henceforth, $\underset{1}{\Omega}$ is no longer arbitrary, and the simple manipulations with moments analogous to those of Sec. V can reduce it only to one of the following elementary forms:

$$i\phi(\bar{x}),\ i(p_k,\phi_{,k}(\bar{x})),\ i(p_k,(p_l,\phi_{,kl}(\bar{x}))),\ldots, \tag{7.2}$$

where the $\phi(\bar{x})$ are always harmonic. It might seem that the manipulations by harmonic $V(\bar{x},t)$ are therefore restricted. However, one has to remember that $\underset{1}{\Omega}$ describes only the most immediate dynamic effects. An exact expression for the exponent Ω in (7.1), known as the *Magnus formula* (also the continuous Baker–Campbell–Hausdorff formula) involves infinite integral series of multiple commutators $[H(t_n),[H(t_{n-1}),\ldots[H(t_2),H(t_1)]\ldots]$. An algorithm to determine its subsequent terms was given[24] and an explicit formal solution was found in Refs. 15 and 16,

$$\Omega = \sum_{n=0}^{+\infty}\epsilon^n\underset{n}{\Omega}, \tag{7.3}$$

$$\underset{n}{\Omega} = (-i)^{n-1}\int_0^\tau\cdots\int_0^\tau L_n(t_n,\ldots,t_1)$$
$$\times H(t_n)\cdots H(t_1)\,dt_n\cdots dt_1, \tag{7.4}$$

where $H(t)$ means the time-dependent Hamiltonian, and the integration kernels are

$$L_n(t_n,\ldots,t_1) = \frac{1}{n}(-1)^{n-1-\Theta_n}\binom{n-1}{\Theta_n}^{-1},$$
$$\Theta_n = \Theta(t_n-t_{n-1})+\cdots+\Theta(t_2-t_1). \tag{7.5}$$

Due to the time dependence of $V(\bar{x},t)$, there may be physical situations when the contribution from $\underset{1}{\Omega}$ vanishes and higher-order terms of (7.3) become dominant. A clear circumstance of this kind occurs for the double loop twice affected by the same potential with inverted signs (see Fig. 12).

The resulting precession effect is then defined by $\underset{2}{\Omega}$ (a generalized method of moments of Sec. V is available; author's notes).

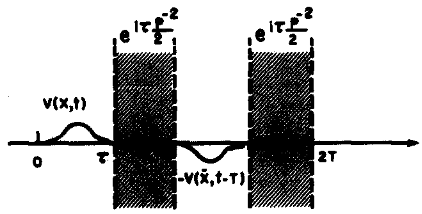


FIG. 12. A double pulse pattern annihilating $\underset{1}{\Omega}$.

A convenient device to check the global algebraic contents of all "precession effects" is the Lie algebra G spanned by $i\bar{p}^2/2$ and by the elements (7.2). Its natural embedding frame is the algebra $\mathscr{A}$ of all "formal expressions"; $\mathscr{A} := \mathscr{F}/\mathscr{I}$, where $\mathscr{F}$ is the complex free algebra of all formal power series of six symbols x_j, p_j $(j=1,2,3)$ with a natural topology, and $\mathscr{I}$ is the ideal generated by the elements: $x_kx_i = x_jx_k, p_kp_j - p_jp_k, x_jp_k - p_kx_j - i\delta_{jk}$ $(j,k=1,2,3)$. Note now the following lemma.

Lemma 7: The smallest Lie algebra containing $i\bar{p}^2/2$ and the imaginary forms (7.2), where $\phi(\bar{x})$ are harmonic polynomials, is dense in $\mathscr{A}$.

Proof: Due to (7.2), G contains any harmonic polynomial $iu(\bar{x})$, as well as any $i(p_k,v_{,k}(\bar{x}))$, where $v(\bar{x})$ is harmonic. This implies

$$[i(p_k,v_{,k}),iu] = iu_{,k}v_{,k} = i\Delta(uv)\in G.$$

Hence, for any harmonic, real polynomials $u_1,\ldots,u_n$ and $v_1,\ldots,v_n$, $i\Delta(u_1v_1+\cdots+u_nv_n)\in G$. However, any polynomial $\omega(\bar{x}) = \Sigma u_j(\bar{x})v_j(\bar{x})$, where $u_j(\bar{x})$ and $v_j(\bar{x})$ are harmonic (author's notes). In turn, any polynomial $f(\bar{x})$ is of the form $f(\bar{x}) = \Delta\omega(\bar{x})$. Hence, every $if(\bar{x})\in G$. The rest of the proof is due to the fact that the Lie algebra containing $i\bar{p}^2/2$ and all polynomials $if(\bar{x})$ must also contain all imaginary polynomials of $\bar{x}, \bar{p}$. ■

This lemma means that there is no algebraic barrier that would prevent one from approximating any unitary operation as a sequence of harmonically induced precession effects.

Focusing operation: Since our results do not include effective prescriptions, the existence of some exact solutions is worth noticing. Consider again the evolution problem in one space dimension. Let $V(q,t) = \gamma(t)q^2/2$ be any loop-creating potential [with $\gamma(t)$ periodic, of period T; the loop period $\tau = nT$]. Then consider the distortion of the loop by $V_1(q,t) = f(t)q$, where $f(t)$ is periodic, of period τ (a periodically varying, homogeneous force). Once more, the evolution operator will split into the cyclic part $U_0(t)$, as in Sec. V, and the "precession part" $W(t)$:

$$\frac{dW}{dt} = -if(t)[U_0(t)^*qU_0(t)]W(t)$$
$$= -if(t)[G(t)q+T(t)p]W(t), \tag{7.6}$$

where $G(t)$ and $T(t)$ are defined by the matrix $u(t)$ of Sec. IV [see (4.4)]. Note now that the time-dependent generators in (7.6) commute to a number. Henceforth, the formula (7.3) and (7.4) for $W(\tau)$ simplifies to

$$W(\tau) \equiv \exp\Big(-i\int_0^\tau f(t)[G(t)q+T(t)p]\,dt\Big), \tag{7.7}$$

where the symbol $\equiv$ means that we have neglected a c-number phase factor. The functions $G(t)$ and $T(t)$ fulfill the differential equations (4.7) and (4.8) with the boundary values

$$\begin{aligned} &G(0)=G(\tau)=1, \quad T(0)=T(\tau)=0, \\ &G'(0)=G'(\tau)=0, \quad T'(0)=T'(\tau)=1, \end{aligned} \tag{7.8}$$

and are linearly independent in $[0,\tau]$. Hence, by choosing suitably the function $f(t)$ in $[0,\tau]$ one can obtain the precession (7.7) in the form of a unitary operator with any linear combination of q and p in the exponent. In particular, taking $f(t)$ such that

$$\int_0^\tau f(t)G(t)dt=0, \quad \int_0^\tau f(t)T(t)dt=a, \tag{7.9}$$

one achieves

$$W(\tau)=e^{-iap}. \tag{7.10}$$

In order to find the $f(t)$ that fulfills (7.9) it is not necessary to solve (4.7) and (4.8) $G(t)$ and $T(t)$. Put

$$f(t)=-a(t/\tau)\gamma(t). \tag{7.11}$$

Then, due to Eqs. (4.7) and (4.8) and the boundary values (7.8) the integral identities (7.9) are immediately fulfilled. The solution (7.10) is exact and consists of a shift of the wave packet along the q axis. Now, consider the corresponding problem in $\mathbb{R}^3$. Let the loop of the "three phase" harmonic potential given by (6.18) be perturbed by $V_1(\bar{x},t)=-a(t/\tau)\gamma(t)x$. The evolution operator within the time $[0,\tau]$ will split into three commuting parts, two of them yielding the identity (the loop condition), while the third one causes the translative precession of the system:

$$U(0,\tau)\equiv\exp\left(-ip_x\int_0^\tau f(t)T(t)dt\right)=\exp(-iap_x). \tag{7.12}$$

Thus, instead of returning to its initial form, Schrödinger's wave packet, after the time τ, is displaced along the x axis, without any change of shape. Should the electron emerge from a hole (or slit) into a space filled with the pulsating field of the perturbed loop, the wave packet should focus again after the time τ at the distance a from its source, and then in the time moments 2τ, 3τ,... at the distances $2a$, $3a$,... (a phenomenon that suggests that perhaps the quality of images in the electronic microscopes might be improved by using time-dependent lense fields).

VIII. THE DOCTOR FAUSTUS' DEVIL OF FINITE PARTICLE SYSTEMS

Some exceptional features of the loop solutions of Secs. III–VII are worth discussion. As operator identities, they represent an arbitrary number of simultaneous (noninteracting) state trajectories. Given a statistical ensemble of Schrödinger's particles in a loop field, the whole ensemble must therefore perform a loop, returning to its initial state, after the loop period τ. A similar ensemble interpretation holds for the free evolution inversion $e^{i\tau\vec{p}^2/2}$. Given a statistical ensemble with improbable initial density distribution (e.g., empty holes inside of densely populated areas), which is then left to evolve freely for an arbitrarily long time, the oscillation patterns of Sec. VI allow us to restore the past ensemble state with all its forgotten details. One might think that such effects are possible only for ensembles of noninteracting particles. However, this seems not to be the case.

The hypothesis about a "hidden order" in a cloud of colliding particles was recently raised by Brewer and Hahn.[19] The behavior of finite particle systems in pulsating fields was investigated by Waniewski[21] who has shown the existence of oscillation patterns generalizing those of Sec. III and producing the operations inverse to the natural evolution of the system (interactions included). The fields used by Waniewski had a disadvantage of not being source-free. However, due to our Lemma 7, it appears that they too have harmonic analogs. If this is the case, we would be on the way to confronting some traditional images associated with thermodynamic irreversibility.

It is a strong conviction of the present-day statistical mechanics that if a system of many interacting particles is let to evolve, the initial state is irreversibly lost in a labyrinth of collisions. It could be recovered only with the help of a mythical entity called the "Maxwell's demon." This entity should perceive the individual particles of the system and act upon them selectively, enforcing their return to the initial, thermodynamically improbable state. On the level of quantum theories, the action of the demon would be additionally blocked by a semantic paradox: each act of perception of a microparticle observable destroys the microparticle state to a point of no return, without providing more reward than a single eigenvalue. Hence, the demon could not act without destroying what he was supposed to restore.

Though the image is suggestive, the argument contains a gap. The traditional statistical mechanics tells only what happens to a large particle system interacting with a relatively unsophisticated surrounding (thermostate, thermic isolator, etc.). It does not tell what might happen to the same system if placed in a sequence of ordered impulses of highly intricate structure, interpretable as an information beam. What happens then, as found in Refs. 19 and 21 and our Sec. VI is the existence of special patterns of the field variations, which can restore the system to its past. These "go back" drifts, in their simplest form, have been known as the spin-eco effects. However, they exist also in infinite-dimensional spaces of states. As they can be induced in blind, they do not engage the familiar antinomy of "destructive knowledge." In practice, there is something else that handicaps them. When the number of system particles grows, the complication of the required operation patterns increases sharply, and so does their sensitivity to the little errors. Henceforth, they might be of no practical importance for the present-day experimental techniques. Yet, they show that there are not one, but two different demonlike entities who oppose the thermodynamical chaos: the Maxwell's demon and the Doctor Faustus' devil. While the first one acts on the principle of detailed insight into the system microstate (with all antinomies involved) the second one plays only a certain universal melody, without worrying about the past, present, and the future of the system.

Note added in proof: Since (4.3) is an antirepresentation, our conclusion about the deconfinement in Sec. IV B, strictly speaking, concerns the behavior of the Heisenberg trajectories in the past. However, it can be seen that it simultaneously implies their deconfinement in the future.

ACKNOWLEDGMENTS

The author is grateful to Dr. Héctor Nava Jaimes and Dr. Arnulfo Zepeda for their kind hospitality at the Departamento de Física del CINVESTAV in Mexico; to Professor Jerzy Plebański, Dr. Magdaleno Medina-Noyola, Dr. Miguel Angel Péres, Dr. José Luis Morán, Dr. Piotr Kielanowski, Dr. Matias Moreno, Dr. William Wassam, and Dr. Kurt Bernardo Wolf for moral support and comments, and to all colleagues at the Departamento de Física for their interest in this paper. Special thanks are due to Dr. Luis M. Villaseñor, Dr. Gerardo Herrera Corral, and Dr. Gabino Torres Vega for the help with the computational aspects of the paper.

The author was supported by COSNET, SEP.

[1]D. Ruelle, "Applications conservant une mesure absolument continue par ropport a dx sur $[0,1]$," Commun. Math. Phys. **55**, 47 (1977).

[2]J. P. Eckmann, "Roads to turbulence in dissipative dynamical systems," Rev. Mod. Phys. **53**, 643 (1981).

[3]J. Hellman, review talk at the Physics Department, CINVESTAV, México, 1982.

[4]R. Haag and D. Kastler, "On algebraic approach to quantum field theory," J. Math. Phys. **5**, 848 (1964).

[5]E. Lubkin, "Theory of multibin test: Definition and existence of extraneous tests," J. Math. Phys. **15**, 663 (1974); "A physical system which can be forced to execute an arbitrary unitary transformation, and its use to perform arbitrary tests," *ibid.* **15**, 673 (1974).

[6]B. Mielnik, "Generalized quantum mechanics," Commun. Math. Phys. **37**, 221 (1974).

[7]R. Haag and U. Bannier, "Comments on Mielnik's generalized (nonlinear) quantum mechanics," Commun. Math. Phys. **60**, 1 (1978).

[8]B. Mielnik, "Mobility of nonlinear systems," J. Math. Phys. **21**, 44 (1980).

[9]B. Mielnik, "Motion and form," in *Current Issues in Quantum Logic* (Plenum, New York, 1981), pp. 465–477.

[10]R. S. Ingarden, private communication, Toruń, November 1981.

[11]B. Mielnik, "Quantum theory without axioms," in *Second Oxford Symposium on Quantum Gravity*, edited by R. Penrose (Oxford U.P., Oxford, 1981).

[12]A. Messiah, *Mecanique Quantique* (Dunod, Paris, 1964).

[13]J. Wilcox, J. Math. Phys. **8**, 962 (1967).

[14]S. Steinberg, "Applications of the Lie Algebraic formulas of Baker, Campbell, Hausdorff, and Zassenhaus," J. Diff. Eqs. **26**, 404 (1977).

[15]I. Biatynicki-Birula, B. Mielnik, and J. Plebański, Ann. Phys. (NY) **51**, 187 (1969).

[16]B. Mielnik and J. Plebański, "Combinatorial approach to Baker–Campbell–Hausdorff exponents," Ann. Inst. H. Poincaré **XII**, 215 (1970).

[17]M. Suzuki, "Decomposition formulas of exponential operators and Lie exponentials with some applications to quantum mechanics and statistical physics," J. Math. Phys. **26**, 601 (1985).

[18]B. Mielnik, "Mobility of Schrödinger's particle," Rep. Math. Phys. **12**, 331 (1977).

[19]R. G. Brewer and E. L. Hahn, "Atomic memory," Sci. Am. **251**(12), 50 (1984).

[20]J. Waniewski, "Theorem about completeness of quantum mechanical motion group," Rep. Math. Phys. **11**, 331 (1977).

[21]J. Waniewski, "Time inversion and mobility of many particle systems," Commun. Math. Phys. **76**, 27 (1980).

[22]J. Plebański, "On the generators of unitary and pseudo-orthogonal groups," report CINVESTAV, Mexico, 1965.

[23]K. B. Wolf, *Integral Transform in Science and Engineering* (Plenum, New York, 1979).

[24]W. Magnus, Commun. Pure Appl. Math. **7**, 649 (1954).

[25]H. F. Trotter, "On the product of semigroups of operators," Proc. Am. Math. Soc. **10**, 545 (1959).

[26]G. Herrera and B. Mielnik, "Manipulations of Schrödinger particle by time dependent quadratic potentials" (in preparation).

Variational processes and stochastic versions of mechanics

J. C. Zambrini[a),b)]

Princeton University, Department of Mathematics, Fine Hall, Washington Road, Princeton, New Jersey 08544

(Received 8 August 1985; accepted for publication 30 April 1986)

The dynamical structure of any reasonable stochastic version of classical mechanics is investigated, including the version created by Nelson [E. Nelson, *Quantum Fluctuations* (Princeton U.P., Princeton, NJ, 1985); Phys. Rev. **150**, 1079 (1966)] for the description of quantum phenomena. Two different theories result from this common structure. One of them is the imaginary time version of Nelson's theory, whose existence was unknown, and yields a radically new probabilistic interpretation of the heat equation. The existence and uniqueness of all the involved stochastic processes is shown under conditions suggested by the variational approach of Yasue [K. Yasue, J. Math. Phys. 22, 1010 (1981)].

I. INTRODUCTION

Nelson's stochastic mechanics is an unorthodox approach of quantum mechanics that attempts to take seriously the probabilistic concepts of this theory. It shows that, in spite of a frequent belief, it is possible to interpret the quantum phenomena in terms of diffusion processes and therefore to use some physical intuition to understand these phenomena.[1–3]

Let us begin by a summary of stochastic mechanics for one particle of unit mass in a scalar potential $V: \mathbb{R}^3 \to \mathbb{R}$. The key assumption of this theory is that the particle performs a diffusion process described by the (Itô) stochastic differential equation

$$dX(t) = b(X(t),t)dt + \sqrt{\hbar}\, dw(t)\,, \tag{1.1}$$

where $b: \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}^3$ is a function with the units of a velocity, called the drift of the process, $\hbar$ is Planck's constant, and $w(t)$ is a Wiener process on $\mathbb{R}^3$. Notice that this kinematical assumption is compatible with the Heisenberg principle since it says that the quadratic variation of the process on a time interval dt is of the order $\hbar\, dt$ or, using the symbolic notation whose meaning is specified by stochastic calculus,[4]

$$(dX(t))^2 = \hbar\, dt\,, \tag{1.2}$$

which is a version of the uncertainty relations.

Now let us fix a time interval $I = [-T/2, T/2]$. It is shown in stochastic mechanics that to every (sufficiently regular) solution ψ of the Schrödinger equation on $\mathbb{R}^3 \times I$,

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2}\Delta\psi + V\psi \tag{1.3}$$

is associated a diffusion process (1.1) with drift

$$b(x,t) = \hbar\, \mathrm{Re}\, \frac{\nabla\psi(x,t)}{\psi(x,t)} + \hbar\, \mathrm{Im}\, \frac{\nabla\psi(x,t)}{\psi(x,t)}\,, \tag{1.4}$$

in such a way that the Born interpretation is valid,

$$P(X(t) \in d^3x) = |\psi(x,t)|^2\, d^3x\,, \quad \forall t \text{ in } [-T/2, T/2]\,, \tag{1.5}$$

for P the probability measure of this process.

Actually, the construction of the processes and the proof of their existence and uniqueness for a given initial (for example) density of probability $|\psi(x, -T/2)|^2\, dx$ is not a simple problem. In particular, at the quantum nodes (zeros of the wave function ψ) the drift (1.4) is manifestly singular, too much anyway for the standard methods of constructing diffusions. Recently, Carlen[5] gave an analytical proof of existence and uniqueness of such a Markovian diffusion for quite general solutions of the Schrödinger equation (1.3) and a given (final) density $|\psi(x,T/2)|^2\, dx$.

After this crucial mathematical result and the publication of Nelson's 1983 course on stochastic mechanics given in Switzerland,[1] one can wonder about the need for further investigations of the conceptual structure of this theory.

Actually, in spite of its successes, stochastic mechanics is a singularity in the field of the physical theories, whose consequences, both from the mathematical and physical points of view, are far from being understood. This is partly due to the unique character of this construction but also to the very implicit form of its key probabilistic concepts. An example of this last aspect is that, given a solution of the Schrödinger equation (1.3), we know everything explicitly about the associated Markovian diffusion ..., except its transition probability, even in very simple cases.

On the other hand, the relevant diffusion processes are "conservative" in the sense of Nelson[1]: their properties are invariant under time reversal. It is well known that the concept of Markovian property is time symmetric, but most of the physical Markovian diffusions are not; in other words, this symmetry is not an intrinsic characteristic of the Markovian diffusion. Nevertheless stochastic mechanics is a fundamentally time symmetric theory since it describes quantum phenomena. One of our key theses is that the class of Markovian processes is not the most natural one to investigate stochastic versions of mechanics.

This natural class of processes was suggested by the variational approach of stochastic mechanics discovered by Yasue.[6–8] He showed that a general diffusion (not necessarily Markovian) is a critical point of the action functional

---

[a)] Address after 30 October 1985: BiBoS Research Center, Fakultat Physik, Bielefeld University, D-4800 Bielefeld 1, Federal Republic of Germany.

[b)] Address after September 1986: University of Warwick, Mathematics Institute, Coventry, England.

$$J[X] = E\left[\int_{-T/2}^{T/2}\left\{\frac{1}{2}L(X(t),DX(t),t) + \frac{1}{2}L(X(t),D_* X(t),t)\right\} dt\right] \tag{1.6}$$

for L the classical Lagrangian of the system [in our case $L(x,\dot{x},t) = \frac{1}{2}|\dot{x}|^2 - V(x,t)$, with $|\cdot|$ the Euclidean norm] if and only if the following Newton equation holds:

$$\tfrac{1}{2}(DD_*X + D_*DX)(t) = -\nabla V. \tag{1.7}$$

In (1.6) and (1.7), DX and D_*X, the two natural generalizations, in stochastic mechanics, of the classical notion of velocity are defined as

$$DX(t) = \lim_{h\downarrow 0} E_t[[X(t+h) - X(t)]/h], \tag{1.8}$$

$$D_*X(t) = \lim_{h\downarrow 0} E_t[[X(t) - X(t-h)]/h], \tag{1.9}$$

where E_t denotes the conditional expectation given $X(t)$, if these limits exists in $L^1(P)$ $\forall t$ in $]-T/2,T/2[$.

To get the result (1.7) one has to assume that the two boundary random variables $X_{-T/2}$ and $X_{T/2}$ are fixed during the variation of the action $J[X]$, in complete analogy with the classical case.

From a probabilistic point of view, however, it is a very unusual hypothesis. These two random variables are not independent, and such a condition seems to involve their joint density $m = m(x,y)$, a probabilistic data used neither in the construction of stochastic mechanics, nor in the usual constructions of Markov processes. In spite of this, the Newton equation (1.7) is indeed an important relation of stochastic mechanics: it specifies the dynamics. Notice, however, that both physical and probabilistic interpretations of the stochastic acceleration are quite obscure in the usual approach of the theory. Why is it not possible to construct a reasonable version of stochastic mechanics using another time symmetric acceleration, for example, $\frac{1}{2}(DDX + D_*D_*X)$?

Other dynamical aspects of stochastic mechanics can be analyzed via a rather formal stochastic calculus of variations whose (1.6) and (1.7) are the basic results.[9]

All this suggests that this variational approach for two fixed end points really has something to do with the probabilistic structure of the involved diffusion processes, even if it seems doubtful that only Markov processes can be constructed from this starting point. But it is worthwhile to observe that several recent developments of stochastic mechanics also suggest that the class of Markov process may be too restrictive for some purposes of this theory.[1]

In this article, we introduce, in a way consistent with the above-mentioned variational starting point, a new class of stochastic processes called "Bernstein processes" associated with a given quantum mechanical evolution in real *and* in imaginary time (i.e., also for the heat equation with potential V). In both cases, only one member of this class is a Markov process: in real time it is the diffusion process of stochastic mechanics, in imaginary time it is a diffusion (without Killing) that yields a radically new probabilistic interpretation of the heat equation.

It turns out that the constraint of two fixed end points, interpreted here as the knowledge of the two quantum densities $|\psi(x,-T/2)|^2 dx$ and $|\psi(x,T/2)|^2 dx$ in real time, and of two nonzero densities $p_{-T/2}(x)dx$ and $p_{T/2}(x)dx$ in imaginary time, is perfectly natural from the time symmetrical point of view of any version of stochastic mechanics and leads to a new perspective on the involved diffusion processes, even for the Markovian representatives.

As a by-product of this construction, we obtain closed formulas for all the transition probabilities of the Markovian representatives. The fact that we know their closed form (unknown till now) will be very useful for the future probabilistic investigations of this theory, and for any practical use of it, in theoretical physics (cf. also the Conclusion).

An illustration of what we expect of this approach is given in the last section (Sec. V), justifying the title of the paper, where we propose a new least action principle in stochastic mechanics and in the imaginary time version of it. This principle is compatible with all the needs of both of these theories and uses nothing but the parabolic equations, which are central in our construction.

The program of this work is the following.

In Sec. III we introduce the two starting kernels of our construction. In imaginary time, it is directly the integral kernel of the Schrödinger semigroup, in real time we show how to associate to a given solution of the Schrödinger equation a parabolic equation of evolution and we start from the kernel of this one. We use these kernels to construct a stochastic process X_t on $I = [-T/2,T/2]$ whose joint density of $X_{-T/2}$ and $X_{T/2}$ is given. The resulting process is generally not Markovian. Only one choice of joint density produces a Markovian diffusion process.

We use in this part an adaptation of a mathematical program proposed a long time ago by Bernstein in order to develop an idea of Schrödinger.[10,11] The program was realized by Jamison and Beurling, with a contribution from Fortet. It is summarized, with some of its key arguments, in the Appendix.

A crucial feature of these processes (the Bernstein processes) is that they are intrinsically time symmetrical. The construction proposed here implies the *a priori* symmetrization of the Markovian representative used by Nelson in stochastic mechanics.

Section IV investigates the (unique) Markovian Bernstein processes associated to the real and imaginary time versions of stochastic mechanics under the two fixed end-points constraint. In the real time situation, it is shown that the result of this construction is indeed the diffusion process of stochastic mechanics. The imaginary time case shows that it is also possible to associate, in a new dynamical way, time reversible diffusion processes (without Killing) to an evolution under the heat equation.

If one chooses invariant boundary densities with nodes (zeros), as we have to for the quantum stationary states, our construction has to be slightly modified because the uniqueness of the process X_t on $I = [-T/2,T/2]$ is lost. Actually, we show that it is sufficient to use the natural decomposition of the state space M into the disjoint domains formed by the nodes to get a unique stationary Bernstein process in each of these domains. The uniqueness of the process on the entire state space follows trivially. Also notice that the same is true in imaginary time because in the stationary situation the two considered starting parabolic equations have, up to the trivial time dependence, the same solutions.

In Sec. V, we characterize in a new variational way the Bernstein processes relevant for the two versions of stochastic mechanics ("variational processes"). Since, in imaginary time, this yields a new probabilistic interpretation of the classical heat equation directly inspired by Schrödinger, the resulting new theory of classical diffusing particles is called "(Schrödinger's) stochastic variational dynamics."

There is no attempt here to find the best regularity conditions for the construction, mainly because our goal is to describe a new constructive frame for any version of stochastic mechanics. Therefore all the conditions given here can be greatly weakened. This work has a twofold motivation. The first is to develop a truly consistent probabilistic extension of the classical variational approach of dynamics and to show that it is more fundamental than previously anticipated in stochastic mechanics since it leads to a completely different probabilistic construction of this theory. The second is to suggest, as in Sec. V, how to use this formulation to discover new conceptual and technical aspects of stochastic mechanics, using the comparison, henceforth possible, between the real and imaginary time dynamics. From our point of view, indeed, the fact that the dynamical structure of stochastic mechanics is shown not to be restricted to the real time description increases notably the reach of this theory.

Finally, since our constructive approach brings to light new interesting physical features, these will be analyzed separately in another publication.[12]

II. SOME CONVENTIONS, NOTATIONS, AND DEFINITIONS

The stochastic processes indexed by I, X_t: $\Omega \to \dot{M}$, $\omega \to \omega(t) \equiv X(t,\omega) \equiv X(t)$ considered here, for I a compact time interval $[-T/2, T/2]$, $\dot{M}$ the one-point compactification of a locally compact separable metric space M, and $\Omega = \Pi_{t\in I}\, \dot{M}$ the compact (under product topology) separable space of the functions ω: $I \to \dot{M}$, are defined on the probability space (Ω, σ_I, P).

The set $\sigma_I = \sigma\{X_t, t\in I\}$ is the Borel sigma algebra of Ω called the natural filtration of X_t and M the state space of the process. For a given sigma algebra $\mathscr{B} = \mathscr{B}(M)$, $(M,\mathscr{B})$ is a measurable space. We will be interested mainly in the case where M is the Euclidean space $\mathbb{R}^n$, or a region Λ of $\mathbb{R}^n$, and $\mathscr{B}$ the associated Borel sigma algebra. When M is not specified, in the following, the result has a more general validity.

Some subsigma algebras of σ_I, or filtrations, will be useful: the past at time t, denoted by $\mathscr{P}_t = \sigma\{X_s, s \leqslant t\}$, the future at time t, $\mathscr{F}_t = \sigma\{X_u, u \geqslant t\}$, and the present at time t, $\mathscr{N}_t = \mathscr{P}_t \cap \mathscr{F}_t$. Other relevant filtrations for the construction are introduced in the Appendix. A process Y_t is said to be $\mathscr{F}_t$ adapted (for example), if, for any t in I, Y_t is $\mathscr{F}_t$ measurable. The (absolute) expectation of an integrable random variable X on (Ω, σ_I, P) will be denoted by $E[X]$ and its conditional expectation with respect to a sigma algebra $\mathscr{B}$ by $E[X \mid \mathscr{B}]$. For the present $\mathscr{N}_t$, the short notation $E_t[X]$ is used and, if Y is another random variable, $E_{x,t}[Y]$ denotes $E[Y \mid X_t = x]$.

If X_t is $\mathscr{P}_t$ adapted, this process is called a $\mathscr{P}_t$ martingale when $E[X_t \mid \mathscr{P}_s] = X_s$ for $s < t$. For example, the usual Wiener process w_t is a $\mathscr{P}_t$ martingale. A $\mathscr{P}_t$ (local) continuous $\mathbb{R}^n$-valued semimartingale X_t admits the decomposition, for $t\in I$,

$$X_t = X_{-T/2} + B_t + M_t\,,$$

with $B_{-T/2} = M_{-T/2} = 0$, where M_t is a continuous $\mathscr{P}_t$ (local) martingale and B_t a continuous $\mathscr{P}_t$-adapted process of bounded variation. This decomposition is unique for X_t and $\mathscr{P}_t$ given and called the Meyer canonical decomposition of the process.

A particularly interesting class of $\mathscr{P}_t$ semimartingales is the following. Assuming that $\underset{t \mapsto X_t}{I \to L^1(P)}$ is continuous, their canonical decomposition is given by

$$X(t) = X\left(\frac{-T}{2}\right) + \int_{-T/2}^{t} DX(s)ds + \sqrt{\hbar}\left\{w(t) - w\left(-\frac{T}{2}\right)\right\},$$

where the forward derivative defined by

$$DX(t) = \lim_{\Delta t \downarrow 0} E\left[\,[X(t+\Delta t) - X(t)]/t \mid \mathscr{P}_t\right] \tag{2.1}$$

is in $L^1(P)$ $\forall t\in I$, $\hbar$ is a positive constant, and $w(t)$ is a $\mathscr{P}_t$-Wiener process. In particular, if X_t satisfies the $\mathscr{P}_t$-(Itô) stochastic differential equation

$$dX(t) = b(X(t),t)dt + \sqrt{\hbar}\, dw(t)\,, \quad t\in I, \quad X_{-T/2} = x\,, \tag{2.2}$$

for b a smooth $\mathbb{R}^n \times I \to \mathbb{R}^n$ function, then $DX(t) = b(X(t),t)$ and the process is a Markovian diffusion process, namely, it has the property that for any t in I, $A\in\mathscr{P}_t$, and $B\in\mathscr{F}_t$, $P(AB \mid \mathscr{N}_t) = P(A \mid \mathscr{N}_t)\cdot P(B \mid \mathscr{N}_t)$. Intuitively, the drift b depends only on the present information, contained in $\mathscr{N}_t$. Notice however that the above-mentioned class of semimartingales is larger: $DX(t)$ is generally past dependent and, in this case, X_t is not Markovian.

An analogous canonical decomposition may be written with respect to the decreasing filtration $\mathscr{F}_t$ and an associated $\mathscr{F}_t$ martingale. In the restrictive (smooth) Markovian case, if one imposes that the same process X_t satisfies the $\mathscr{F}_t$-Itô stochastic differential equation

$$dX(t) = b_*(X(t),t)dt + \sqrt{\hbar}\, dw_*(t)\,, \quad t\in I, \quad X_{T/2} = y\,, \tag{2.3}$$

then using the (general) backward derivative

$$D_*X(t) = \lim_{\Delta t \downarrow 0} E\left[\,[X(t) - X(t-\Delta t)]/t \mid \mathscr{F}_t\right], \tag{2.4}$$

we get $D_*X(t) = b_*(X(t),t)$ for $w_*(t)$ an $\mathscr{F}_t$ martingale. We will also use, in the same Markovian case, for any C^∞ function, f: $M \times I \to M$ the forward and backward derivatives

$$Df(X(t),t) = \left(\frac{\partial}{\partial t} + b(X(t),t)\cdot\nabla + \frac{\hbar}{2}\Delta\right) f(X(t),t) \tag{2.5}$$

and

$$D_* f(X(t),t) = \left(\frac{\partial}{\partial t} + b_*(X(t),t)\cdot\nabla - \frac{\hbar}{2}\Delta\right) f(X(t),t)\,. \tag{2.6}$$

These formulas follow from a Taylor expansion of f and the use of the given $\mathscr{P}_t$ and $\mathscr{F}_t$ stochastic differential equations. Actually they are valid in a wider context. For example, if the underlying diffusion process is not Markovian, and $dX(t) = b(t)dt + \sqrt{\hbar}\,dw(t)$, where $b(t)$ is simply $\mathscr{P}_t$-adapted and $\int_I |b(t)|dt < \infty$ a.s., they are still true. Their conditions of validity are the ones of Itô's formula.[4]

Finally, most of the time, under an integral sign, we will denote simply by dx the volume element $dx_1 \cdots dx_n$, when $M = \mathbb{R}^n$.

A number of equations, in this article, hold only almost everywhere, but it will be clear from the context.

III. BERNSTEIN PROCESSES IN TERMS OF THE SCHRÖDINGER AND HEAT EQUATIONS

Consider the following equation of $M \times I = \mathbb{R}^n \times [-T/2, T/2]$:

$$\sigma\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2}\Delta\psi + V\psi\,, \tag{3.1}$$

with some initial condition $\psi(x,-T/2) = \psi_{-T/2}(x)$, where $\hbar$ is a given positive constant and V: $M \to \mathbb{R}$ a smooth (real) potential. If $\sigma = -1$, this is the heat equation, if $\sigma = i$, this is the Schrödinger equation. It is well known that the heat equation, which is parabolic, is easy to analyze in probabilistic terms, in contrast with the Schrödinger equation. Nevertheless suppose that we know a (sufficiently regular) solution of this last equation in $L^2(M)$,

$$\psi(x,t) \equiv e^{(R+iS)(x,t)/\hbar}\,. \tag{3.2}$$

Then we will associate to this particular quantum dynamics another parabolic equation on $M \times I$,

$$-\hbar\frac{\partial\varphi^*}{\partial t} = -\frac{\hbar^2}{2}\Delta\varphi^* + \vartheta\varphi^*\,, \tag{3.3}$$

where the, generally time-dependent, modified potential ϑ is defined in terms of this solution of Schrödinger and of the physical potential V by

$$\vartheta = (\nabla R)^2 + \hbar\Delta R - V\,. \tag{3.4}$$

From now on, we will parallel as far as possible the construction of time reversible stochastic processes associated to the two starting equations, using directly (3.1) when $\sigma = -1$ and via (3.3) when $\sigma = i$. In spite of these apparently unrelated starting points, both cases will correspond to the realization of an essentially unique stochastic dynamical structure. Therefore, for the sake of symmetry, the two considered parabolic equations are, respectively, rewritten as

$$-\hbar\frac{\partial\theta^*}{\partial t} = -\frac{\hbar^2}{2}\Delta\theta^* + V\theta^* \equiv H\theta^*\,, \tag{3.5}$$

$$-\hbar\frac{\partial\varphi^*}{\partial t} = -\frac{\hbar^2}{2}\Delta\varphi^* + \vartheta\varphi^* \equiv \mathscr{H}\varphi^* \tag{3.6}$$

[note (3.6)≡(3.3)]. All the functions considered in the construction will be real. Since our motivation comes from theoretical physics, we will refer to the heat equation (3.5) as an imaginary time Schrödinger equation. In this context, it would be preferable to denote the parameters of the Schrödinger equation and of Eq. (3.5) in a different way, but we will conform here with the physicist's custom.

We need to impose some regularity conditions on the potentials of Eqs. (3.5) and (3.6) sufficient to assure, in particular, that the kernels of the semigroups associated to (3.5) and (3.6) are strictly positive.

Theorem 3.1: Let V be a real-valued function, Hölder continuous almost everywhere on $M = \mathbb{R}^n$ bounded from below and such that the Hamiltonian $H = -(\hbar^2/2)\Delta + V$ is essentially self-adjoint. Then the fundamental solution $h(s,x,t,y)$ of the parabolic equation (3.5) satisfies the following properties.

For a given initial condition $\theta^*_{-T/2}(x)$ in the space $B(M)$ of the continuous bounded functions on M equipped with the norm $\|\theta^*\| = \sup_{x\in M}|\theta^*(x)|$ the classical solution of (3.5) [(3.6)] on $M \times I$ is given by

$$(T_{t+T/2}\theta^*_{-T/2}) = \int_M \theta^*_{-T/2}(x)\,h\left(-\frac{T}{2},x,t,y\right)dx\,,$$

where T_t: $B(M) \to B(M)$ is a homogeneous and strongly continuous contraction semigroup whose infinitesimal generator is H. Here T_t is called the Schrödinger semigroup. The domain $\mathscr{D}(H)$ is defined as the set

$$\{g \in B(M) \text{ such that s-}\lim_{t\downarrow 0}[(T_t - 1)/t]g \text{ exists}\}.$$

Moreover, (a) $h(s,x,t,y) \equiv h(x,t-s,y)$ can be chosen to be jointly continuous in x, y, and $(t-s)$; (b) $h(s,x,t,y)$ is strictly positive, and $h(s,x,t,y) \underset{|y|\to\infty}{\longrightarrow} 0$;

$$\text{(c)}\ \lim_{\Delta s\downarrow 0}\int_{S_\epsilon(x)} h(s,x,s+\Delta s,y)\,dy = 1,$$

for $S_\epsilon(x)$ the sphere of center x and radius ϵ;

$$\text{(d1)}\ \lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\int_M (y-x)h(s,x,s+\Delta s,y)\,dy = 0\,;$$

$$\text{(d2)}\ \lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\int_M (y-x)^2 h(s,x,s+\Delta s,y)\,dy = \hbar I\,,$$

where, in dimension >1, $(y-x)^2$ means $(y-x) \otimes (y-x)$ and I is an identity matrix on M; (d3) there is a $\delta > 0$ such that

$$\int_M |y-x|^{2+\delta}h(s,x,s+\Delta s,y)\,dy = o(\Delta s)\,;$$

$$\text{(e)}\ \lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\left[1 - \int_M h(s,x,s+\Delta s,y)\,dy\right] = V(x)\,;$$

and (f) in (d1) and (d2) the region of integration M can be replaced by $S_\epsilon(x)$. The analog is true for the integral kernel $k(s,x,t,y)$ of the semigroup U_t associated to Eq. (3.6), and the modified potential ϑ. We sketch only some of these classical arguments in this case.

The integral representation of the solution is well known. Notice that the Hölder continuity of ϑ is necessary for the existence and continuity of $\partial\theta^*/\partial t$ and $\Delta\theta^*$.

(b) Since ϑ is real and bounded from below, the strict positivity of k follows from the Feynman–Kac (integral) representation of the semigroup U_t, valid under our hypothesis,

$$k(s,x,t,y) = E_{y,t}\left[\exp - \int_s^t \vartheta(w(\tau),\tau)\,d\tau\right] q(s,x,t,y)\,,$$

where q is the transition identity of the n-dimensional Wiener process $w(\tau)$, namely the (strictly positive) fundamental solution for the heat equation (on $M = \mathbb{R}^n$),

$$q(s,x,t,y) = (2\pi(t-s))^{-n/2}\exp(-|y-x|^2/2(t-s))$$

and $E_{y,t}$ is the conditional expectation of the Wiener process $w(\tau)$, knowing $w(t) = y$.

(a) The joint continuity of k follows from the continuity of ϑ.

Choosing $(s,x) = (0,0)$ for notational simplicity, we check that (c) is true for a bounded modified potential $|\vartheta(x,\tau)| \leqslant \alpha$. The general case of a bounded below potential ϑ is obtained via a sequence of such potentials,

$$\vartheta_n(x,\tau) = \begin{cases} \vartheta(x,\tau)\,, & \text{if } |\vartheta(x,\tau)| \leqslant n\,, \\ n\,, & \text{otherwise}\,, \end{cases} \qquad n\in\mathbb{N}\,,$$

such that $\vartheta(x,\tau) = \lim_{n\to\infty} \vartheta_n(x,\tau)$.

It was shown by Kac[13] that, in these conditions, the function $K(y,t) = k(0,0,y,t)$ satisfies the integral equation $(M = \mathbb{R}^n)$

$$K(y,t) + \int_0^t \int_M q(\tau,x,t,y)\vartheta(x,\tau)K(x,\tau)\,dx\,d\tau$$
$$= q(0,0,y,t)\,.$$

Now, by (b) and the bound on ϑ,

$$K(x,\tau) \leqslant \exp(\alpha\tau)e^{-x^2/2\tau}/(2\pi\tau)^{n/2}\,,$$

and therefore, using this in the integral equation

$$\int_{S_\epsilon(0)} K(y,t)\,dy$$
$$\leqslant \int_{S_\epsilon(0)} \int_0^t \int_M q(\tau,x,t,y)\alpha e^{\alpha\tau}\frac{e^{-x^2/2\tau}}{(2\pi\tau)^{n/2}}\,dx\,d\tau\,dy$$
$$+ \int_{S_\epsilon(0)} q(0,0,y,t)\,dy\,.$$

The first term reduces to

$$\alpha\left(\frac{e^{\alpha t}}{\alpha} - \frac{1}{\alpha}\right)\int_{S_\epsilon(0)} \frac{e^{-y^2/2t}}{(2\pi t)^n}\,dy$$

and then vanishes at $\lim_{t\downarrow 0}$. Since

$$\lim_{t\downarrow 0} \int_{S_\epsilon(0)} q(0,0,y,t)\,dy = 1\,,$$

the property (c) is satisfied by the kernel k.

For (d3) we consider

$$\lim_{t\downarrow 0} \int_M |y-x|^4 k(0,x,t,y)\,dy\,.$$

The right-hand side of the integral equation shows that we need to evaluate

$$\int_M |y-x|^4 q(0,x,t,y)\,dy\,.$$

This is $3t^{2n}$ and therefore vanishes at $\lim_{t\downarrow 0}$. For the second term of the right-hand side in the integral equation we observe that

$$\left|\int_M |y-x|^4 \int_0^t \int_M q(\tau,x,t,y)\vartheta(x,\tau)K(x,\tau)\,dy\right|$$
$$\leqslant \int_M |y-x|^4 \int_0^t \int_M q(\tau,x,t,y)\alpha e^{\alpha\tau}\frac{e^{-x^2/2\tau}}{(2\pi\tau)^{n/2}}\,dx\,d\tau\,dy$$
$$-\alpha\left(\frac{e^{\alpha t}}{\alpha} - \frac{1}{\alpha}\right)\int_M |y-x|^4\frac{e^{-y^2/2t}}{(2\pi t)^{n/2}}\,dy\,.$$

Since this is also zero at $\lim_{t\downarrow 0}$, (d3) is true for $\delta = 2$.

Properties (d1), (d2), and (e) are immediate consequences from the fact that (3.6) may be interpreted as the diffusion equation for a stochastic process with zero drift, diffusion matrix $\hbar I$ and Killing rate ϑ. In (d1) and (d2) the region of integration is all of M, but taking into account the relation (d3) it may be replaced by the sphere $S_\epsilon(x)$ of center x and radius ϵ. This is the content of (f). For example,

$$0 = \lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\int_M (y-x)k(s,x,s+\Delta s,y)\,dy$$
$$= \lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\int_{S_\epsilon(x)} (y-x)k(s,x,s+\Delta s,y)\,dy$$
$$+ \lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\int_{\overline{S}_\epsilon(x)} (y-x)k(s,x,s+\Delta s,y)\,dy\,,$$

where $\overline{S}_\epsilon(x)$ is the complement of the sphere. But since, by (d3) for $\alpha < 2+\delta$,

$$\lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\int_{\overline{S}_\epsilon(x)} |y-x|^\alpha k(s,x,s+\Delta s,y)\,dy$$
$$\leqslant \frac{1}{\epsilon^{2+\delta-\alpha}}\lim_{\Delta t\downarrow 0}\frac{1}{\Delta t}$$
$$\times \int_{\overline{S}_\epsilon(x)} |y-x|^{2+\delta}k(s,x,s+\Delta s,y)\,dy = 0\,,$$

(d1) can be replaced by the "truncated rate" on $S_\epsilon(x)$.

Remark 1: All the results of the Theorem 3.1 are also true if $\theta^*_{-T/2(x)}$ $[\varphi^*_{-T/2}(x)]$ are continuous functions in $L^2(M)$, namely if the semigroups T_t, U_t are considered on this Hilbert space, and, actually, under much more general conditions.[14]

In imaginary time, the hypothesis V bounded below is natural since this function is given. The analogous condition on ϑ in real time, that is, for a given *solution* of the Schrödinger equation, seems hard to control since it involves the given physical potential V and this particular solution. Moreover, the explicit form of ϑ [Eq. (3.4)] suggests that our condition is almost never satisfied. This is not the case.

Lemma 3.1.1: Let $\psi(x,t) = e^{(R+iS)(x,t)/\hbar}$ be a given regular solution of the starting Schrödinger equation (3.1) (for $\sigma = i$). If the potential V and $\partial S/\partial t$ are bounded from below, ϑ is bounded from below.

Proof: The substitution of $\psi = e^{(R+iS)/h}$ in the Schrödinger equation yields the coupled nonlinear partial differential equations on $M \times I$

$$\frac{\partial R}{\partial t} = -\nabla R \cdot \nabla S - \frac{\hbar}{2} S \nabla S , \tag{3.7}$$

$$\frac{\partial S}{\partial t} = -\frac{1}{2}(\nabla S)^2 + \frac{1}{2}(\nabla R)^2 + \frac{\hbar}{2}\Delta R - V . \tag{3.8}$$

In particular, in using Eq. (3.8), the modified potential ϑ reduces to

$$\vartheta = V + 2\frac{\partial S}{\partial t} + (\nabla S)^2 , \tag{3.9}$$

and then the conclusion holds. □

From the physical point of view, the constraint $\partial_t S$ bounded from below is a natural sufficient condition in the following sense. If this is true and $\rho(x,t) \equiv |\psi(x,t)|^2$, then

$$\int_M \partial_t S \rho \, dx > -\infty .$$

But using Eq. (8), this is

$$\int_M \left\{ -\frac{1}{2}(\nabla S)^2 + \frac{1}{2}(\nabla R)^2 + \frac{h}{2}\Delta R - V \right\} \rho \, dx > -\infty ,$$

and since $\rho(x,t) = e^{2R(x,t)}$, it may be seen that

$$\int_M \frac{\hbar}{2}\Delta R \cdot \rho \, dx = \int_M -(\nabla R)^2 \rho \, dx .$$

Therefore our constraint simplifies to

$$\int_M \left\{ \frac{1}{2}(\nabla S)^2 + \frac{1}{2}(\nabla R)^2 + V \right\} \rho \, dx < \infty . \tag{3.10}$$

This is a finite energy condition. In particular,

$$\int_M \left\{ \frac{1}{2}(\nabla S)^2 + \frac{1}{2}(\nabla R)^2 \right\} \rho \, dx \equiv \int_M |\nabla \psi|^2 \, dx < \infty \tag{3.11}$$

is independent of time and therefore is true on $I = [-T/2, T/2]$ if $\|\nabla \psi_{-T/2}\|_2^2 < \infty$. Under this condition, Carlen proved the existence of the processes of stochastic mechanics along the constructive lines proposed originally by Nelson.[5]

Now, assuming that we are in the conditions of the Theorem 3.1, let us define the strictly positive functions of six variables for $-T/2 \leqslant s < t \leqslant T/2$ and x, y, z in the state space $(M, \mathscr{B})$,

$$h(s,x;t,y;u,z) = \frac{h(s,x,t,y)h(t,y,u,z)}{h(s,x,u,z)} , \tag{3.12}$$

and for $A \in \mathscr{B}$,

$$H(s,x;t,A;u,z) = \int_A h(s,x;t,y;u,z) \, dy \equiv \int_A h(s,x;t,dy;u,z) ,$$

and similarly for the kernel of the semigroup U_t,

$$k(s,x;t,y,u,z) = \frac{k(s,x,t,y)k(t,y,u,z)}{k(s,x,u,z)} ,$$

$$K(s,x;t,A;u,z) = \int_A k(s,x,t,dy;u,z) . \tag{3.13}$$

Such a function, for example K, has three important properties.

(K1) $\forall x, z \in M$ and $-T/2 \leqslant s < t < u \leqslant T/2$, the mapping $A \mapsto K(s,x;t,A,u,z)$ is a probability measure on $\mathscr{B}$, since, by the semigroup property of U_t, i.e., the Chapman–Kolmogorov equation,

$$\int_M k(s,x;t,dy;u,z) = \frac{1}{k(s,x,u,z)} \int_M k(s,x,t,dy)k(t,y,u,z) = 1 .$$

(K2) For any fixed A in $\mathscr{B}$, the mapping $(x,z) \mapsto K(s,x;t,A;u,z)$ is $\mathscr{B} \times \mathscr{B}$ measurable since the image of (x,z) is a product of continuous functions.

(K3) For any A, B in $\mathscr{B}$, $-T/2 \leqslant s < t < u < v \leqslant T/2$,

$$\int_B K(s,w;t,A;u,y)K(s,w;u,dy;v,z) = \int_A K(s,w;t,dx;v,z)K(t,x;u,B;v,z) .$$

It is evident in using the above-mentioned definition of K. Actually, since (K3) is valid $\forall A, B \in \mathscr{B}$, it says simply that

$$\text{(K3}') \; k(s,w;t,x;u,y) \cdot k(s,w;u,y;v,z) = k(s,w;t,x;v,z)k(t,x;u,y;v,z) .$$

These three properties define a reciprocal transition probability function K in the sense of Jamison.[15]

Interpreting this function K as an analogy of the transition of probability for a Markov process, one can construct a stochastic process X_t, $t \in I = [-T/2, T/2]$ for which a joint probability of $X_{-T/2}$ and $X_{T/2}$ will play the role of the initial distribution. The story of this idea is very interesting and quite old. It is therefore surprising that it was never exploited in mathematical physics. We summarize this story, and the relevant characteristics of such processes, the Bernstein processes, in the Appendix and we use the same conventions. For the moment, we simply observe that, by definition, X_u is a Bernstein process on I if, for any bounded Borel measurable g, $-T/2 \leqslant t < u < v \leqslant T/2$,

$$E[g(X_u) | \mathscr{P}_t \cup \mathscr{F}_v] = E[g(X_u) | X_t, X_v] ,$$

for $\mathscr{P}_t$ the past at time t and $\mathscr{F}_v$ the future at time v. In general, such a process is not Markovian.

In the next theorem the function K may be replaced by $H = H(s,x;t,B;u,y)$.

Theorem 3.2: Let M a locally compact metric space, $K = K(s,x;t,B;u,y)$ as before, and $m = m(x,y)$ a probability measure on $\mathscr{B} \times \mathscr{B}$. Then there is a unique probability measure P_m such that with respect to (Ω, σ_I, P_m), X_t, $t \in I$, is a Bernstein process and

(1) $P_m(X_{-T/2} \in B_s, X_{T/2} \in B_E) = m(B_S \times B_E)$,

where B_S, $B_E \in \mathscr{B}$ (where we use S for "start," E for "end"); and

(2) $\forall -T/2 \leqslant s < t < u \leqslant T/2$, $B \in \mathscr{B}$,

$$P_m(X_t \in B | X_s, X_u) = K(s, X_s; t, B; u, X_u) .$$

Furthermore the probability $P_m(C)$ of the cylinder event

$$C = \{X_{-T/2} \in B_S , X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n , X_{T/2} \in B_E\}$$

is given by

(3) $\displaystyle\int_{B_S\times B_E} dm(x,y)\int_{B_1} K\left(-\frac{T}{2},x;t_1,dx_1;\frac{T}{2},y\right)$

$$\times\int_{B_2} K\left(t_1,x_1;t_2,dx_2;\frac{T}{2},y\right)$$

$$\cdots\int_{B_n} K\left(t_{n-1},x_{n-1};t_n,dx_n;\frac{T}{2},y\right)$$

or equivalently by

$(3)_*$ $\displaystyle\int_{B_S\times B_E} dm(x,y)\int_{B_1} K\left(-\frac{T}{2},x;t_1,dx_1;t_2,x_2\right)$

$$\times\int_{B_2} K\left(-\frac{T}{2},x;t_2,dx_n;t_3,x_3\right)$$

$$\cdots\int_{B_n} K\left(-\frac{T}{2},x;t_n,dx_n;\frac{T}{2},y\right).$$

In other words, the finite-dimensional distribution density of X_t is given by

$$\rho_m(dx_1,t_1,dx_2,t_2,\ldots,dx_n,t_n)$$
$$=\int_{M\times M} dm(x,y)k\left(-\frac{T}{2},x;t_1,dx_1;\frac{T}{2},y\right)$$
$$\cdots k\left(t_{n-1};x_{n-1};t_n,dx_n;\frac{T}{2},y\right),$$

or, equivalently, by

$$\int_{M\times M} dm(x,y)k\left(-\frac{T}{2},x;t_1,dx_1;t_2,x_2\right)$$
$$\cdots k\left(-\frac{T}{2},x;t_n,dx_n;\frac{T}{2},y\right).$$

The proof of this theorem is given in the Appendix.

In Jamison's construction of Bernstein processes[14] one shows how to construct these processes starting from a given Markov process. From the point of view of a physical dynamics, it is not the relevant way. The choice of our starting kernels h and k is precisely dictated by a dynamical point of view (cf. Secs. IV and V). However, most of the critical steps of his original constructions may be adapted to this new purpose.

For example, the next theorem shows that there is only one possible choice of joint density $m=m(x,y)$ for $X_{-T/2}$ and $X_{T/2}$ such that the Bernstein process constructed in Theorem 3.2 is a Markov process.

Theorem 3.3: Let $(M,\mathscr{B})$ be the state space, $k=k(s,x,t,y)$ the strictly positive $\mathscr{B}\times\mathscr{B}$ (x,y)-measurable kernel of the given semigroup U_t, $K=K(s,x,t,A,u,y)$ as in Theorem 3.2, $m=m(x,y)$ a probability measure on $\mathscr{B}\times\mathscr{B}$, and X_t, $t\in I$, the Bernstein process for this m. Then X_t, $t\in I$, is a Markovian diffusion process $\Leftrightarrow$ there are two real and nonzero bounded functions of the same sign on M, $\varphi^*_{-T/2}$ and $\varphi_{T/2}$ such that

$$m(B_S\times B_E)$$
$$=\int_{B_S\times B_E}\varphi^*_{-T/2}(x)k\left(-\frac{T}{2},x,\frac{T}{2},y\right)$$
$$\times\varphi_{T/2}(y)\,dx\,dy\,,\quad B_S,B_E\ \text{in}\ \mathscr{B}\,.$$

Remarks: (1) The analogous result is true for the process Z_t associated to the Schrödinger semigroup T_t.

(2) In Sec. IV we shall consider mechanical stationary states for which this result will be used on connected domains Λ of the original state space M and with appropriate boundary conditions for the kernels k. On some of these domains, $\varphi^*_{-T/2}$ and $\varphi_{T/2}$ both will be negative.

(3) No constructive way to find $\varphi^*_{-T/2}$ and $\varphi_{T/2}$ is given here (cf. Theorem 3.4).

Proof: We give the proof for the process in real time X_t.

(i) $\Leftarrow$ the substitution of this particular m in the finite-dimensional distribution density of X_t given in Theorem 3.2 yields, after simplifications,

$$\rho_m(dx_1,t_1,dx_2,t_2,\ldots,dx_n t_n)$$
$$=\int_{B_S\times B_E}\varphi^*_{-T/2}(x)k\left(-\frac{T}{2},dx,t_1,dx_1\right)$$
$$\cdots k(t_{n-1},dx_{n-1},t_n,dx_n)$$
$$\times k\left(t_n,x_n,\frac{T}{2},dy\right)\varphi_{T/2}(y)\,. \tag{3.14}$$

On the other hand, let us define on $M\times I$ the backward evolution of $\varphi_{T/2}$,

$$\varphi(x,s)\equiv\int_M k\left(s,x,\frac{T}{2},y\right)\varphi_{T/2}(y)\,dy\,. \tag{3.15}$$

Notice that, for $-T/2\leqslant s<t\leqslant T/2$, we also have

$$\varphi(x,s)=\int_M k(s,x,t,z)\varphi(z,t)\,dz\,. \tag{3.16}$$

Indeed, by the semigroup property of k,

$$\int_M k\left(s,x,\frac{T}{2},y\right)\varphi_{T/2}(y)\,dy$$
$$=\int_M\int_M k(s,x,t,z)k\left(t,z,\frac{T}{2},y\right)dz\,\varphi_{T/2}(y)\,dy$$
$$=\int_M k(s,x,t,z)\varphi(z,t)\,dz\,.$$

Let us also define for $B\in\mathscr{B}$, $-T/2\leqslant s<t\leqslant T/2$ and $x,y\in M$,

$$P(s,x,t,B)=\frac{1}{\varphi(x,s)}\int_B k(s,x,t,y)\varphi(y,t)\,dy$$

and the associated density

$$p(s,x,t,y)=k(s,x,t,y)[\varphi(y,t)/\varphi(x,s)]\,. \tag{3.17}$$

We claim that P is the (forward) transition probability of a Markov process. Indeed, p is a non-negative-valued function by hypothesis on k and on $\varphi_{T/2}$.

Moreover $P(s,x,t,\cdot)$ is a probability on $\mathscr{B}$, since, using (3.15) and (3.17),

$$\int_M p(s,x,t,y)\,dy=\frac{1}{\varphi(x,s)}\int_M k(s,x,t,y)\varphi(y,t)\,dy=1\,.$$

Notice that, since k is strictly positive, the result of the backward evolution of $\varphi_{T/2}$, given by (3.15), has the sign of $\varphi_{T/2}$. (This is true, as a matter of fact, even if $\varphi_{T/2}$ has zeros, but we shall consider this case separately.)

Now $P(s,\cdot,t,B)$ is $\mathscr{B}$-measurable for fixed $s<t$, $B\in\mathscr{B}$, since p is a product of continuous functions.

The Chapman–Kolmogorov equation for P is a direct consequence of the semigroup property of k.

Finally, given that k is an integral kernel, P is normal in the sense of Dynkin,[16]

$$P(s,x,s,B)=\chi_B(x)\,,$$

for any s in I, B in $\mathscr{B}$.

Therefore P has all the properties of a transition probability. On the other hand, using (3.15) for the integration with respect to y in (3.14), and after substitution of the transition densities p given by (3.17), the finite-dimensional distribution of X_t reduces to

$$\begin{aligned}\rho_m&(dx_1,t_1,dx_2,t_2,\ldots,dx_n,t_n)\\ &=\int_M\rho\Big(dx\,,-\frac{T}{2}\Big)P\Big(-\frac{T}{2},x,t_1,dx_1\Big)P(t_1,x_1,t_2,dx_2)\\ &\quad\cdots P(t_{n-1},x_{n-1},t_n,dx_n)\end{aligned}\tag{3.18}$$

for the initial distribution

$$\rho\Big(x,-\frac{T}{2}\Big)\,dx\equiv\varphi^*_{-T/2}(x)\varphi\Big(x,-\frac{T}{2}\Big)\,dx\,.\tag{3.19}$$

This is nothing but the finite-dimensional distribution for a Markov process X_t with transition probability $P(s,x,t,B)$ and initial distribution $\rho(dx,-T/2)\equiv\rho(x,-T/2)dx$. Consequently, a version of the Bernstein process constructed with the particular joint density m of the hypothesis is indeed Markovian.

In the same way, starting from the definition on $M\times I$,

$$\begin{aligned}\varphi^*(y,t)&\equiv\int_M\varphi^*_{-T/2}(x)k\Big(-\frac{T}{2},x,t,y\Big)\,dx\,,\\ &\qquad -T/2<t<T/2\,,\quad x,y\in M\,,\end{aligned}\tag{3.20}$$

one verifies that for $A\in\mathscr{B}$, $-T/2\leqslant s<t\leqslant T/2$,

$$P_*(s,A,t,y)=\frac{1}{\varphi^*(y,t)}\int_A\varphi^*(x,s)k(s,x,t,y)\,dx$$

is the backward transition of a Markov process, with density

$$p_*(s,x,t,y)=[\varphi^*(x,s)/\varphi^*(y,t)]k(s,x,t,y)\,.\tag{3.21}$$

This enables us to express also the finite-dimensional distribution of X_t as

$$\begin{aligned}\rho_m&(dx_1,t_1,dx_2,t_2,\ldots,dx_n,t_n)\\ &=\int_M P_*(t_1,dx_1,t_2,x_2)P_*(t_2,dx_2,t_3,x_3)\\ &\quad\cdots P_*\Big(t_n,dx_n,\frac{T}{2},y\Big)\rho\Big(dy\,,\frac{T}{2}\Big)\,,\end{aligned}\tag{3.22}$$

for the final distribution

$$\rho\Big(y\,,\frac{T}{2}\Big)\,dy\equiv\varphi^*\Big(y\,,\frac{T}{2}\Big)\,\varphi_{T/2}(y)\,dy\,.\tag{3.23}$$

A necessary condition that P is indeed the transition probability of a diffusion process is the fulfillment of

$$\lim_{\Delta s\downarrow 0}\int_{S_\epsilon(x)}p(s,x,s+\Delta s,y)\,dy=0\,,\quad\forall\epsilon>0\,,\quad\forall x\in M\,.$$

In order to evaluate this Lindeberg-type expression, we introduce Eq. (3.17),

$$\lim_{\Delta s\downarrow 0}\int_{S_\epsilon(x)}k(s,x,s+\Delta s,y)\,\frac{\varphi(y,s+\Delta s)}{\varphi(x,s)}\,dy\,.$$

In the hypothesis on ϑ of the Theorem 3.1, φ is of class C^2. If we assume, for simplicity, that its derivatives are bounded there is a z in $]x,y[$ and τ in $]s,s+h[$ such that

$$\begin{aligned}\varphi(&y,s+\Delta s)\\ &=\varphi(x,s)+(y-x)\nabla\varphi(x,s)+\Delta s\,\frac{\partial\varphi}{\partial s}(x,s)\\ &\quad+\frac{1}{2}(y-x)^2\nabla^2\varphi(z,s)+\frac{1}{2}\,\Delta s^2\,\frac{\partial^2\varphi}{\partial s^2}(x,\tau)\,,\end{aligned}$$

so we need to evaluate

$$\begin{aligned}&\lim_{\Delta s\downarrow 0}\int_{S_\epsilon(x)}k(s,x,s+\Delta s,y)\,dy+\frac{\nabla\varphi^*(x,s)}{\varphi^*(x,s)}\\ &\quad\times\lim_{\Delta s\downarrow 0}\int_{S_\epsilon(x)}(y-x)k(s,x,s+\Delta s,y)\,dy\\ &\quad+\frac{1}{\varphi(x,s)}\lim_{\Delta s\downarrow 0}\int_{S_\epsilon(x)}\frac{1}{2}(y-x)^2\nabla^2\\ &\quad\times\varphi(z,s)k(s,x,s+\Delta s,y)\,dy\,.\end{aligned}\tag{3.24}$$

The third integral is bounded by

$$\max_{y\in S_\epsilon(x)}|\nabla^2\varphi(z,s)|\lim_{\Delta s\downarrow 0}\int_{S_\epsilon(x)}\frac{1}{2}(y-x)^2k(s,x,s+\Delta s,y)\,dy$$

and for $\delta>0$ this is

$$\begin{aligned}&\leqslant\max_{y\in S_\epsilon(x)}|\nabla^2\varphi(z,s)|\lim_{\Delta s\downarrow 0}\frac{1}{2\epsilon^\delta}\\ &\quad\times\int_{S_\epsilon(x)}|y-x|^{2+\delta}k(s,x,s+\Delta s,y)\,dy\\ &\leqslant\max_{y\in S_\epsilon(x)}|\nabla^2\varphi(x,s)|\lim_{\Delta s\downarrow 0}\frac{1}{2\epsilon^\delta}\\ &\quad\times\int_M|y-x|^{2+\delta}k(s,x,s+h,y)\,dy\,.\end{aligned}$$

By (d3) of Theorem 3.1 this is zero. Moreover the same condition shows that

$$\begin{aligned}&\lim_{\Delta s\downarrow 0}\int_{S_\epsilon(x)}k(s,x,s+\Delta s,y)\,dy\\ &\quad\leqslant\lim_{\Delta s\downarrow 0}\frac{1}{\epsilon^{2+\delta}}\int_{S_\epsilon(x)}|y-x|^{2+\delta}k(s,x,s+\Delta s,y)\,dy\\ &\quad\leqslant\frac{1}{\epsilon^{2+\delta}}\lim_{\Delta s\downarrow 0}\int_M|y-x|^{2+\delta}k(s,x,s+\Delta s,y)\,dy=0\end{aligned}$$

and

$$\begin{aligned}&\lim_{\Delta s\downarrow 0}\int_{S_\epsilon(x)}(y-x)k(s,x,s+\Delta s,y)\,dy\\ &\quad\leqslant\lim_{\Delta s\downarrow 0}\frac{1}{\epsilon^{1+\delta}}\int_{S_\epsilon(x)}|y-x|^{2+\delta}k(s,x,s+\Delta s,y)\,dy\\ &\quad\leqslant\frac{1}{\epsilon^{1+\delta}}\lim_{\Delta s\downarrow 0}\int_M|y-x|^{2+\delta}k(s,s,s+\Delta s,y)\,dy=0\,,\end{aligned}$$

therefore the two other integrals of (3.24) are zero and for any x in M, s, in I, and $\epsilon>0$,

$$\lim_{\Delta s\downarrow 0}\int_{S_\epsilon(x)}p(s,x,s+\Delta s,y)\,dy=0\,.\tag{3.25}$$

This is called the stochastic continuity of the transition probability.[16] In particular, this implies that $P(s,x,t,B)$ is uniquely defined by the generator of the associated Markov transition semigroup. Other important characteristics of the Markovian diffusion process constructed here are given in Corollaries 3.3.1 and 3.3.2.

(ii) $\Rightarrow$ Let $P(s,x,t,B)$ [resp. $P_*(s,B,t,y)$] be the forward (backward) transition probability of the Markovian process X_t, $p(s,x,t,y)$ [$p_*(s,x,t,y)$] its forward (backward) density and $\rho(dy,t) = \rho(y,t)dy$ the density of probability at time t.

It follows from the relation of duality between the forward and backward transition probabilities relative to the density,

$$P\left(s,x,\frac{T}{2},B_E\right) = \frac{1}{\rho(x,t)} \int_{B_E} P_*\left(s,x,\frac{T}{2},y\right) \rho\left(dy,\frac{T}{2}\right),$$

that $P(s,x,T/2,\cdot)$ is absolutely continuous with respect to $\rho(\cdot,T/2)$. Let us denote by $\beta(x,\cdot)$ its density. Similarly, for the starting time $t = -T/2$, we denote by $\gamma(x,\cdot)$ the density of $P(-T/2,x,T/2,\cdot)$ with respect to $\rho(\cdot,T/2)$. Finally, $\alpha(x,\cdot)$ will be the density of $P(s,x,t,\cdot)$ with respect to the Lebesgue measure $d(\cdot)$.

Now, as a Markov process, X_t satisfies

$$\begin{aligned}
&P_m(X_{-T/2}\in B_S,\ X_t\in B,\ X_{T/2}\in B_E) \\
&\quad = \int_{B_S} \rho\left(dx,-\frac{T}{2}\right) \int_B P\left(-\frac{T}{2},x,t,dz\right) \\
&\qquad \times \int_{B_E} P\left(t,z,\frac{T}{2},dy\right) = \int_{B_S} \rho\left(dx,-\frac{T}{2}\right) \\
&\qquad \times \int_B \alpha(x,z)dz \int_{B_E} \beta(z,y)\, \rho\left(dy,\frac{T}{2}\right).
\end{aligned}$$

On the other hand, as a Bernstein process with joint density m, X_t also satisfies, according to Theorem 3.2,

$$\begin{aligned}
&P_m(X_{-T/2}\in B_S,\ X_t\in B,\ X_{T/2}\in B_E) \\
&\quad = \iint_{B_S\times B_E} dm(x,y) \int_B K\left(-\frac{T}{2},x,t,dz,,\frac{T}{2},y\right) \\
&\quad = \int_{B_S} \rho\left(dx,-\frac{T}{2}\right) \int_{B_E} \gamma(x,y)\, \rho\left(dy,\frac{T}{2}\right) \\
&\qquad \times \int_B k\left(-\frac{T}{2},x,t,z,\frac{T}{2},y\right) dz .
\end{aligned}$$

The comparison between the two expressions of P_m shows that for almost all x,y (with respect to $\rho_{-T/2}\times\rho_{T/2}$),

$$\gamma(x,y)k\left(-\frac{T}{2},x,t,z,\frac{T}{2},y\right) = \alpha(x,z)\beta(z,y),$$

which means that, by definition of k,

$$\gamma(x,y) = \frac{\alpha(x,z)}{k(-T/2,x,t,z)}\, k\left(-\frac{T}{2},x,\frac{T}{2},y\right) \frac{\beta(z,y)}{k(t,z,T/2,y)}$$

and therefore

$$\begin{aligned}
&m(B_S\times B_E) \\
&\quad = \int_{B_S} \rho\left(dx,-\frac{T}{2}\right) \int_{B_E} P\left(-\frac{T}{2},x,\frac{T}{2},y\right) \\
&\quad = \int_{B_S} \rho\left(dx,-\frac{T}{2}\right) \int_{B_E} \gamma(x,y)\rho\left(dy,\frac{T}{2}\right) \\
&\quad = \int_{B_S\times B_E} \frac{\rho(x,-T/2)\alpha(x,z)}{k(-T/2,x,t,z)} \\
&\qquad \times k\left(-\frac{T}{2},x,\frac{T}{2},y\right) \frac{\beta(z,y)\rho(y,T/2)}{k(t,z,T/2,y)}\, dx\, dy ,
\end{aligned}$$

which is indeed of the expected form. □

Remark: The absolute continuity of $m(B_S\times B_E)$ with respect to $\rho_{-T/2}\times\rho_{T/2}$ for any couple of time $(-T/2, T/2)$ is a sufficient condition for the simultaneous existence of the forward and backward transitions P and P_*.

It will be useful to summarize the results on the Markovian Bernstein process (we shall soon see that it is unique in Theorem 3.4) in two corollaries.

Corollary 3.3.1: Let $k = k(s,x,t,y)$ be the strictly positive kernel of the starting semigroup U_t. Then the Markovian Bernstein process X_t of the Theorem 3.3 is characterized by the following properties.

(1) $\forall t\in I$, $z\in M$, up to a normalization, its probability density ρ is given by

$$\rho(z,t) = \varphi^*(z,t)\varphi(z,t),$$

where φ^* and φ are defined according to Eqs. (3.15) and (3.20) by

$$\varphi^*(z,t) = \int_M \varphi^*_{-T/2}(x)k\left(-\frac{T}{2},x,t,z\right) dx$$

and

$$\varphi(z,t) = \int_M k\left(z,t,y,\frac{T}{2}\right)\varphi_{T/2}(y)\, dy .$$

(2) The densities of its forward and backward transition probabilities satisfy, for $x,y\in M$, $-T/2\leqslant s<t\leqslant T/2$,

$$p(s,x,t,y) = k(s,x,t,y)[\varphi(y,t)/\varphi(x,s)]$$

and

$$p_*(s,x,t,y) = [\varphi^*(x,s)/\varphi^*(y,t)]k(s,x,t,y) .$$

(3) The following relation of duality between the forward and backward densities of transition is valid, for $x,y\in M$, $-T/2\leqslant s<t\leqslant T/2$:

$$\rho(x,s)p(s,x,t,y) = p_*(s,x,t,y)\rho(y,t) .$$

(4) The finite-dimensional distribution of the process X_t may also be written as

$$\begin{aligned}
&\rho_m(dx_1,t_1,dx_2,t_2,\ldots,dx_n,t_n) \\
&\quad = P_*(t_1,dx_1,t_2,x_2)\cdots P_*(t_{i-1},dx_{i-1},t_i,x_i) \\
&\qquad \times \rho(dx_i,t_i)\cdot P(t_i,x_i,t_{i+1},dx_{i+1}) \\
&\qquad \cdots P(t_{n-1},x_{n-1},t_n,dx_n) .
\end{aligned}$$

Remark: The analogous result is true for the imaginary time case. We shall denote by

$$Q(s,x,t,B) = \int_B q(s,x,t,dy)$$

$$\left(Q_*(s,A,t,y) = \int_A q_*(s,dx,t,y)\right)$$

the forward (backward) transition probability of the Markovian Bernstein process Z_t and its associated density, and by $P(dx,t) = p(x,t)dx$ its distribution of probability.

Corollary 3.3.2: If $\varphi_{T/2}, \varphi^*_{-T/2}$ are positive, of class C^2, and have bounded first and second derivatives, the Markovian Bernstein process X_t is a diffusion process whose forward and backward drifts are the $(M = \mathbb{R}^n)$-valued functions b and b_* given by

$$b(x,t) = \hbar(\nabla\varphi/\varphi)(x,t)\,,$$

$$b_*(x,t) = -\hbar(\nabla\varphi^*/\varphi^*)(x,t)\,,$$

and the diffusion matrix the matrix-valued function C and C_* (same notations)

$$C(x,t) = C_*(x,t) = \hbar I\,,$$

for I the $n\times n$ identity matrix.

Proof of Corollary 3.3.1: We consider only the process in real time X_t.

(1) According to (3.14), (3.15), and (3.20) we find that the density $\rho_m(dx_1,t_1) \equiv \rho(dx_1,t_1)$ is, for $x_1 \in M$, $t_1 \in I$,

$$\rho_m(dx_1,t_1) = \int_M dx\, dy\, \varphi^*_{-T/2}(x)k\left(x,-\frac{T}{2}dx_1,t_1\right)k\left(x_1,t_1,y,\frac{T}{2}\right)\varphi_{T/2}(y)$$

$$= \int_M \varphi^*_{-T/2}(x)k\left(x,-\frac{T}{2},dx,t_1\right)dx \int_M k\left(x_1,t_1,y,\frac{T}{2}\right)\varphi_{T/2}(y)\,dy = \varphi^*(x_1,t_1)\varphi(x_1,t_1)\,.$$

We shall see later that this product may indeed be normalized as a probability density (Theorem 4.4).

(2) These results have been found in the proof of the Theorem 3.3: Eqs. (3.17) and (3.21).

(3) The relation of duality is an immediate consequence of the explicit form of the transition densities p and p_* and the density ρ.

(4) In introducing the above-mentioned p, p_*, and ρ in the given finite-dimensional ρ_m we obtain Eq. (3.14).

Proof of the Corollary 3.3.2: By the assumptions about $\varphi_{T/2}$ and $\varphi^*_{-T/2}$ and the definitions (3.15) and (3.20) for the backward and forward evolutions, $\varphi(x,s)$ and $\varphi^*(y,t)$ have the same properties. In particular, the following Taylor expansion makes sense, for some z in $]x,y[$, τ in $]s,s+\Delta s[$:

$$\varphi(y,s+\Delta s) = \varphi(x,s) + (y-x)\nabla\phi(x,s) + \Delta s\frac{\partial\varphi}{\partial s}(x,s) + \frac{1}{2}(y-x)^2\nabla^2\varphi(z,s) + \frac{1}{2}(\Delta s)^2\frac{\partial^2\varphi}{\partial s^2}(x,\tau)\,.$$

Using this in the density of the forward transition probability given in Corollary 3.3.1 (2) (and denoting $\partial\varphi/\partial s$ by $\dot\varphi$) it follows from the definition of the forward drift for a diffusion process that

$$b(x,s) = \lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\int_{S_\epsilon(x)}(y-x)p(s,x,s+\Delta s,y)\,dy = \lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\int_{S_\epsilon(x)}(y-x)\frac{\varphi(y,s+\Delta s)}{\varphi(x,s)}k(s,x,s+\Delta s,y)\,dy$$

$$= \lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\int_{S_\epsilon(x)}(y-x)\left[1+\frac{\nabla\varphi}{\varphi}(x,s)(y-x)+\frac{\dot\varphi}{\varphi}(x,s)\cdot\Delta s+\frac{1}{2}\frac{\nabla^2\varphi(z,s)}{\varphi(x,s)}(y-x)^2\right.$$
$$\left.+\frac{1}{2}\frac{\ddot\varphi(x,\tau)}{\varphi(x,s)}(\Delta s)^2\right]k(s,x,s+\Delta s,y)\,dy$$

$$= \lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\int_{S_\epsilon(x)}(y-x)k(s,x,s+\Delta s,y)\,dy + \frac{\nabla\varphi}{\varphi}(x,s)\lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\int_{S_\epsilon(x)}(y-x)^2k(s,x,s+\Delta s,y)\,dy$$
$$+\frac{\dot\varphi}{\varphi}(x,s)\lim_{\Delta s\downarrow 0}\int_{S_\epsilon(x)}(y-x)k(s,x,s+\Delta s,y)\,dy + \frac{1}{2\varphi(x,s)}\lim_{\Delta s\downarrow 0}\int_{S_\epsilon(x)}(y-x)^3\nabla^2\varphi(z,s)k(s,x,s+\Delta s,y)\,dy$$
$$+\frac{\ddot\varphi(x,\tau)}{2\varphi(x,s)}\lim_{\Delta s\downarrow 0}\Delta s\int_{S_\epsilon(x)}(y-x)k(s,x,s+\Delta s,y)\,dy\,.$$

The first error integral is bounded by

$$\max_{y\in S_\epsilon(x)}|\nabla^2\varphi(z,s)|\lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\int_{S_\epsilon(x)}|y-x|^3k(s,x,s+\Delta s,y)\,dy$$

(or the integral on all of M). By the Theorem 3.1 (d3), this term is zero, as is the other error integral. Since we are in the conditions where the "infinitesimal rates" (d1) and (d2) of the same theorem are identical to the truncated rates on $S_\epsilon(x)$, we get indeed

$$b(x,s) = \hbar(\nabla\varphi/\varphi)(x,s)\,.$$

The same computation for the backward drift,

$$b_*(y,t) = \lim_{\Delta t\downarrow 0}\frac{1}{\Delta t}\int_{S_\epsilon(y)}(y-x)\,p_*(t-\Delta t,x,t,y)\,dx\,,$$

using the density of the backward transition probability [Corollary 3.3.1 (2)], yields

$$b_*(y,t) = -\hbar(\nabla\varphi^*/\varphi^*)(y,t)\,.$$

On the other hand, the diffusion matrix C of the process is defined by

$$C(x,s)=\lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\int_{S_\epsilon(x)}(y-x)^2 p(s,x,s+\Delta s,y)\,dy\,.$$

It follows from the same Taylor expansion as before that

$$\begin{aligned}C(x,s)&=\lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\int_{S_\epsilon(x)}(y-x)^2 k(s,x,s+\Delta s,y)\,dy+\frac{\nabla\varphi}{\varphi}(x,s)\lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\int_{S_\epsilon(x)}(y-x)^3 k(s,x,s+\Delta s,y)\,dy\\&\quad+\frac{\dot\varphi}{\varphi}(x,s)\lim_{\Delta s\downarrow 0}\int_{S_\epsilon(x)}(y-x)^2 k(s,x,s+\Delta s,y)\,dy+\frac{1}{2\varphi(x,s)}\lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\\&\quad\times\int_{S_\epsilon(x)}(y-x)^4\nabla^2\varphi(z,s)k(s,x,s+\Delta s,y)\,dy+\frac{\ddot\varphi(x,\tau)}{2\varphi(x,s)}\lim_{\Delta s\downarrow 0}\Delta s\int_{S_\epsilon(x)}(y-x)^2 k(s,x,s+\Delta s,y)\,dy\,.\end{aligned}$$

The first error integral is bounded by

$$\max_{y\in S_\epsilon(x)}|\nabla^2\varphi(z,s)|\lim_{\Delta s\downarrow 0}\frac{1}{\Delta s}\int_{S_\epsilon(x)}(y-x)^4 k(s,x,s+\Delta s,y)\,dy$$

and so is zero by the Theorem 3.1 (d3). The other error integral vanishes trivially. It follows from (d2) (same theorem) that, if I is an identity matrix, then

$$C(x,t)=\hbar I\,.$$

The (identical) backward diffusion matrix is found in the same way. Of course, all the results concerning the imaginary time process Z_t are obtained along the same lines. □

The point of Bernstein processes is that they are constructed from the data of two boundary probabilities. Now, according to Theorem 3.3, we have strong constraints on the two functions $\varphi^*_{-T/2}$ and $\varphi_{T/2}$ associated to the Markovian representative of the Bernstein process, and therefore on the joint probability $m=m(x,y)$ of the boundary random variables $X_{-T/2}$ and $X_{T/2}$. On the other hand, the two marginals of this joint probability are, by definition, the initial and final distributions $\rho(dx,-T/2)$ and $\rho(dy,T/2)$ of the Markovian Bernstein process we are looking for.

By the Corollary 3.3.1 (1) these conditions on the marginals of m can be expressed in terms of the given kernel and the boundary density probabilities as, in real time,

$$\begin{aligned}&\varphi^*_{-T/2}(x)\int_M k\Big(x,-\frac{T}{2},y,\frac{T}{2}\Big)\varphi_{T/2}(y)\,dy=\rho\Big(x,-\frac{T}{2}\Big),\\&\varphi_{T/2}(y)\int_M\varphi^*_{-T/2}(x)k\Big(x,-\frac{T}{2},y,\frac{T}{2}\Big)\,dx=\rho\Big(y,\frac{T}{2}\Big),\end{aligned}\tag{3.26}$$

and the analogous system in imaginary time.

Now observe that, from the beginning of this construction, the pair of functions $\varphi^*_{-T/2}$, $\varphi_{T/2}$ was never really specified. Our data are the kernel k and the two boundary densities of probability $\rho(x,-T/2)$ and $\rho(y,T/2)$. This means that Eq. (3.26) constitutes a system of nonlinear functional equations for $\varphi^*_{-T/2}$, $\varphi_{T/2}$.

We will call this system the Schrödinger system. Indeed, it was derived more than fifty years ago by Schrödinger for the Gaussian kernel and represents the solution of his construction of time symmetrical diffusion processes.[17]

The problem of existence and uniqueness of the solution for the Schrödinger system is not at all a trivial one.

It was investigated successively by Bernstein,[11] Fortet,[18] and Beurling.[19] This last author proved that if k is bounded away from zero and infinity, $0<a<k<b<\infty$, then the system of Schrödinger has one and only one pair of positive (actually of same signs) solutions $\varphi^*_{-T/2}$ and $\varphi_{T/2}$.

Jamison found a way to extend Beurling's proof without this restriction on k (see Ref. 15).

We recall the following definition: A measure μ on $(M,\mathscr{B})$ is σ-finite if there is a sequence c_n in M such that $\cup_n c_n=M$ and $\mu(c_n)<\infty$.

Theorem 3.4: Let $\rho(dx,-T/2)$ and $\rho(dy,T/2)$ be two strictly positive probability measures on $\mathscr{B}$, for $(M,\mathscr{B})$ a state space whose M is a locally compact separable metric space. Let $k(-T/2,x,T/2,y)=k(x,y)$ be a given kernel, everywhere continuous and strictly positive on $M\times M$. Then there is a unique pair (m,π) of measures on $\mathscr{B}\times\mathscr{B}$, where m is a probability measure and π a σ-finite product measure, such that

$$\begin{aligned}dm(x,y)&=k(x,y)d\pi(x,y)\\&=k(x,y)\varphi^*_{-T/2}(x)\varphi_{T/2}(y)dx\,dy\end{aligned}$$

and the conditions on the marginals of the measure m are given by the Schrödinger system (3.26).

Remark: In Sec. IV we will use this result not on all of the original state space M, but on connected domains Λ of M, as mentioned in Remark (1) of Theorem 3.3.

This is the end of the preparatory work for our construction of the two possible versions of stochastic mechanics. At this point we do not have even a physical kinematics; we did not introduce a notion of velocity, for example. [The forward and backward derivatives (2.1) and (2.4) of Sec. II are two reasonable candidates, but what about an acceleration?] What we did is to describe a method to construct Bernstein processes from the kernels of two semigroups associated to imaginary time and real time Schrödinger equations. In order to understand the relation of this construction with physical dynamics, we have to describe in a much more explicit way the (unique) Markovian representative of these classes of Bernstein processes.

IV. THE MARKOVIAN BERNSTEIN PROCESS AND STOCHASTIC MECHANICS

In this section, we consider independently the two realizations of Markovian Bernstein processes.

A. Imaginary time stochastic mechanics or (Schrödinger's) stochastic variational dynamics

For two arbitrarily chosen strictly positive densities of probability $p_{-T/2}(x)$ and $p_{T/2}(y)$, the Schrödinger system takes the form ($M=\mathbb{R}^n$)

$$\theta^*_{-T/2}(x)\int_M h\left(x,-\frac{T}{2},y,\frac{T}{2}\right)\theta_{T/2}(y)\,dy=p_{-T/2}(x)\,,$$

$$\theta_{T/2}(y)\int_M \theta^*_{-T/2}(x)h\left(x,-\frac{T}{2},y,\frac{T}{2}\right)dx=p_{T/2}(y)\,, \tag{4.1}$$

for h the fundamental solution of the heat equation (3.5).

Since we are in the conditions of the Theorem 3.4, we have existence and uniqueness of the solution $\theta^*_{-T/2}$ and $\theta_{T/2}$ for the Schrödinger system (4.1). It follows from Corollaries 3.3.1 and 3.3.2 that the unique Markovian Bernstein process is completely characterized, since we know $\theta^*(z,t)$ and $\theta(z,t)$ on all of $M\times[-T/2,T/2]$, namely

$$\theta^*(z,t)=\int_M\theta^*_{-T/2}(x)h\left(-\frac{T}{2},x,t,z\right)dx \tag{4.2}$$

and

$$\theta(z,t)=\int_M h\left(z,t,y,\frac{T}{2}\right)\theta_{T/2}(y)\,dy\,. \tag{4.3}$$

Notice that indeed $\theta^*(z,t)$ satisfies (3.5), and $\theta(z,t)$ satisfies the "adjoint" equation (on $M\times I$) under time reversal, namely

$$\hbar\frac{\partial\theta}{\partial t}=-\frac{\hbar^2}{2}\Delta\theta+V\theta\,, \tag{4.4}$$

with the final condition $\theta_{T/2}(y)$.

According to the Corollaries 3.3.1 and 3.3.2 we know everything about the associated Bernstein process, but let us approach the properties of this process in a more "dynamical" way.

Proposition 4.1: In the conditions of Theorem 3.4, the probability density p of the Markovian Bernstein process Z_t is a weak solution on $M\times I$ of

$$\frac{\partial p}{\partial t}+\operatorname{div}J=0\,, \tag{4.5}$$

where the probability current J is defined by the $M\times I\to M$ function

$$J(x,t)=(\hbar/2)[\theta^*\nabla\theta-\theta\nabla\theta^*](x,t) \tag{4.6}$$

and $\theta^*(x,t)$, $\theta(x,t)$ follows from (4.2) and (4.3).

Proof: For any $f\in C_0^2(M)$, using Corollary 3.3.1 and Eqs. (3.5) and (4.4),

$$\begin{aligned}\frac{d}{dt}\int_M f(x)p(x,t)dx&=\frac{d}{dt}\int_M f(x)\theta^*(x,t)\theta(x,t)\,dx\\&=\int_M f\left[\frac{\hbar}{2}\Delta\theta^*-\frac{V}{\hbar}\theta^*\right]\theta\\&\quad+f\theta^*\left[-\frac{\hbar}{2}\Delta\theta+\frac{V}{\hbar}\theta\right]dx\,,\end{aligned}$$

and after integration by parts

$$=\int_M\frac{\hbar}{2}[\theta^*\nabla\theta-\theta\nabla\theta^*]\nabla f\,dx\,.$$

Since the test function f is arbitrary in $C_0^2(M)$, Eq. (5) is satisfied. □

Remark: The equation of continuity (4.5) is the differential form of a (global) conservation of probability. In particular, we always assume that p has been normalized as a probability density.

The form of Eq. (4.5) justifies the following.

Definition: The current velocity of the Markovian Bernstein process Z_t is the $M\times I\to M$ function $\overline{V}$ such that

$$J=p\overline{V}\,. \tag{4.7}$$

Taking into account (4.6), the current velocity reduces to

$$\overline{V}=\hbar\nabla\log(\theta/\theta^*)^{1/2}\,. \tag{4.8}$$

Since $\overline{V}$ is a gradient, it will be useful to introduce $\overline{S}=\hbar\log(\theta/\theta^*)^{1/2}$ such that

$$\overline{V}=\nabla\overline{S}\,. \tag{4.9}$$

Definition: The osmotic velocity of the Markovian Bernstein process Z_t is the $M\times I\to M$ function $\overline{U}$ such that

$$\overline{U}=\hbar\log p^{1/2}\,. \tag{4.10}$$

In other words, using Corollary 3.3.1 (1),

$$\overline{U}=\hbar\nabla\log(\theta^*\theta)^{1/2}\,, \tag{4.11}$$

and it is natural to introduce $\overline{R}=\hbar\log(\theta^*\theta)^{1/2}$ such that

$$\overline{U}=\nabla\overline{R}\,.$$

Now it may be seen that the forward and backward drifts B and B_* of the Markovian Bernstein process Z_t, given by Corollary 3.3.2, may be expressed in terms of the current and osmotic velocity as

$$B(x,t)=(\overline{V}+\overline{U})(x,t)=\hbar(\nabla\theta/\theta)(x,t) \tag{4.12}$$

and

$$B_*(x,t)=(\overline{V}-\overline{U})(x,t)=-\hbar(\nabla\theta^*/\theta^*)(x,t)\,. \tag{4.13}$$

Since we know the equations of motion of θ and θ^*, namely

$$\hbar\frac{\partial\theta}{\partial t}=\frac{\hbar^2}{2}\Delta\theta+V\theta \tag{4.14}$$

and

$$-\hbar\frac{\partial\theta^*}{\partial t}=-\frac{\hbar^2}{2}\Delta\theta^*+V\theta^*\,, \tag{4.14*}$$

it is easy to find that the dynamics of the current and osmotic velocities $\overline{U}$ and $\overline{V}$ is contained in the coupled nonlinear partial differential system on $M\times I$ (assuming V of class C^1),

$$\frac{\partial \overline{U}}{\partial t} = -\frac{\hbar}{2}\,\text{grad div}\,\overline{V} - \text{grad}\,\overline{V}\cdot\overline{U}\,,$$
$$\frac{\partial \overline{V}}{\partial t} = -\frac{\hbar}{2}\,\Delta\overline{U} - \overline{U}\,\nabla\overline{U} - \overline{V}\,\nabla\overline{V} + \nabla V. \tag{4.15}$$

A priori, if we forget the origin of this system of equations, it is a complicated matter to show the existence and uniqueness of its solutions. Here, it is sufficient to observe that the change of variables

$$\theta^* \equiv e^{(\overline{R}-\overline{S})/\hbar}, \tag{4.16}$$

with the above-mentioned $\overline{R}$ and $\overline{S}$, linearizes (4.15) since, by construction, θ^* is the solution of the heat equation (4.14)*. The same remark is valid for $\theta \equiv e^{(\overline{R}+\overline{S})/\hbar}$, the solution of Eq. (4.14).

In summary, for any chosen pair of positive densities $p_{-T/2}(x)$ and $p_{T/2}(y)$, we can in principle [if we are able to solve the Schrödinger system (4.1)] find $\theta^*_{-T/2}$ and $\theta_{T/2}$ such that the forward and backward evolutions (4.2) and (4.3) enable us to construct a Markovian Bernstein diffusion process Z_t on $I = [-T/2, T/2]$, with drifts given by (4.12) and (4.13) on $M \times I$ and the chosen boundary probability densities. This time-symmetric process Z_t is, in this way, naturally associated to the starting heat equation (i.e., the imaginary time Schrödinger equation).

Since the time interval $I = [-T/2, T/2]$ is arbitrary, we have indeed a construction valid for all times.

However, this construction is manifestly not sufficient for any interesting boundary densities of probability.

To see this, let us consider a stationary situation.

If we choose the two boundary densities equal to an invariant density of the form

$$p_j(x) = \phi_j^2(x)\,, \tag{4.17}$$

for $\phi_j(x)$ in $L^2(M,dx)$, a real eigenfunction of the Hamiltonian H in (3.5), distinct from the ground state, then these probabilities have zeros ("nodes") and we cannot directly use our previous results (cf. in particular Theorem 3.4).

As we said before, our constructive approach is still valid in this case, but we need to modify the starting kernel h to take into account the boundary conditions created by the presence of the nodes.

Since this situation is natural in real time, namely for the excited states of the Schrödinger equation, we postpone the discussion of this aspect until the third part of this section.

Before concluding the discussion of the imaginary time process Z_t, we have to specify in which sense the system of partial differential equations (4.15) indeed describes a dynamics. The answer is contained in the following proposition.

Proposition 4.2: In the conditions of the Theorem 3.4, and assuming that V is of class C^1, Eq. (4.15) is equivalent to the Newton equation

$$\tfrac{1}{2}(DDZ + D_*D_*Z)(t) = \nabla V. \tag{4.18}$$

Proof: This is a simple computation. Due to the definitions (2.1) and (2.4) of the forward and backward derivatives, and the formulas (4.12) and (4.13) for the drifts of Z_t we get

$$DZ = \overline{V} + \overline{U}$$

and

$$D_*Z = \overline{V} - \overline{U}.$$

Now by (2.5),

$$\begin{aligned} DDZ &= D\overline{V} + D\overline{U} \\ &= \frac{\partial \overline{V}}{\partial t} + (\overline{V}+\overline{U})\Delta\overline{V} + \frac{\hbar}{2}\Delta\overline{V} + \frac{\partial \overline{U}}{\partial t} \\ &\quad + (\overline{V}+\overline{U})\Delta\overline{U} + \frac{\hbar}{2}\Delta\overline{U}, \end{aligned}$$

and by (2.6),

$$\begin{aligned} D_*D_*Z &= D_*\overline{V} - D_*\overline{U} \\ &= \frac{\partial \overline{V}}{\partial t} + (\overline{V}-\overline{U})\nabla\overline{V} - \frac{\hbar}{2}\Delta\overline{V} - \frac{\partial \overline{U}}{\partial t} \\ &\quad - (\overline{V}-\overline{U})\nabla\overline{U} + \frac{\hbar}{2}\Delta\overline{U}. \end{aligned}$$

Therefore

$$\frac{1}{2}(DDZ + D_*D_*Z) = \frac{\partial \overline{V}}{\partial t} + \overline{V}\cdot\nabla\overline{V} + \overline{U}\cdot\nabla\overline{U} + \frac{\hbar}{2}\Delta\overline{U},$$

and Eq. (4.18) is indeed modified to Eq. (4.15). □

Proposition 4.2 justifies partially the claim that we have constructed a stochastic version of mechanics. It will be confirmed by the results of Sec. V. Notice that the left-hand side of Eq. (4.18) defines the natural notion of acceleration for this construction. It is time symmetric and reduces to the classical acceleration if $t \to Z(t)$ has a differential strong derivative $DZ = dZ/dt$ in $L^1(P)$. Also observe that the sign of the right-hand side of Eq. (4.18) is "wrong" with respect to the classical Newton equation, but correct if we interpret the time parameter t as an "imaginary time $i\tau$." This is why, from the beginning, the starting parabolic equation (3.5) was interpreted as an imaginary time Schrödinger equation. The sense of this remark will be clearer in Sec. IV B. This dynamical theory of diffusion processes associated to a new probabilistic interpretation of the classical heat equation (3.5) will be called hereafter "(Schrödinger's) stochastic variational dynamics" cf. also Sec. V.

B. Real time stochastic mechanics

Due to the form of the *given* solution (3.2) for the Schrödinger equation on $M \times I$, the two quantum boundary densities of probability to consider for the Schrödinger system reduce to

$$\rho(x, -T/2) = e^{2R_{-T/2}(x)/\hbar} \tag{4.19}$$

and

$$\rho(y, T/2) = e^{2R_{T/2}(y)/\hbar} \tag{4.20}$$

when

$$R_{-T/2}(x) \equiv R(x, -T/2)\,, \quad R_{T/2}(y) = R(y, T/2)\,,$$

and therefore this Schrödinger system (3.27) is modified to (for $M = \mathbb{R}^n$)

$$\varphi^*_{-T/2}(x)\int_M k\left(x,-\frac{T}{2},y,\frac{T}{2}\right)\varphi_{T/2}(y)dy=e^{2R_{-T/2}(x)/\hbar},$$

$$\varphi_{T/2}(y)\int_M \varphi^*_{-T/2}(x)k\left(x,-\frac{T}{2},y,\frac{T}{2}\right)dx=e^{2R_{T/2}(y)/\hbar},\tag{4.21}$$

for k the fundamental solution of the parabolic equation (3.6).

Suppose first that we are in the conditions of the (Theorem 3.1 and) Theorem 3.4. Then we have existence and uniqueness of the solutions $\varphi^*_{-T/2}$ and $\varphi_{T/2}$, but it is remarkable that the explicit solution of this problem can be found.

Theorem 4.3: If $\psi(\cdot,t)=e^{(R+iS)(\cdot,t)/\hbar}$ is a continuous solution in $L^2(M)$ of the Schrödinger equation (3.1) (for $\sigma=i$) and if the two quantum boundary densities are free of zeros, the unique solution of the Schrödinger system (4.21) is given by the two continuous functions

$$\varphi^*_{-T/2}(x)=e^{(R_{-T/2}-S_{-T/2})(x)/\hbar}$$

and

$$\varphi_{T/2}(y)=e^{(R_{T/2}+S_{T/2})(y)/\hbar}.$$

Proof: According to Theorem 3.1, the forward propagation of $\varphi^*_{-T/2}$ is (using the notations introduced in Theorem 3.3)

$$\varphi^*(y,t)=\int_M \varphi^*_{-T/2}(x)k\left(-\frac{T}{2},x,t,y\right)dx,$$

$$-T/2<t<T/2,\quad x,y\in M,\tag{4.22}$$

where k is the fundamental solution of the parabolic equation (3.6),

$$-\hbar\frac{\partial\varphi^*}{\partial t}=-\frac{\hbar^2}{2}\Delta\varphi^*+\vartheta\varphi^*.\tag{4.23}$$

Notice that, although the modified potential ϑ [defined in (3.4)] is generally time dependent, it is, like the physical potential $V=V(x)$, invariant under time reversal since $R\to R$ under time inversion. Consequently, the equation "adjoint" to (4.23) under time reversal is simply

$$\hbar\frac{\partial\varphi}{\partial t}=-\frac{\hbar^2}{2}\Delta\varphi+\vartheta\varphi,\tag{4.24}$$

for some final condition in $\varphi_{T/2}(x)$. Therefore, the same kernel k determines the backward propagation [by (3.15)],

$$\varphi(x,t)=\int_M k\left(x,t,y,\frac{T}{2}\right)\varphi_{T/2}(y)\,dy.\tag{4.25}$$

On the other hand, according to the Lemma 3.1.1, the substitution of $\psi(x,t)=e^{(R+iS)(x,t)/\hbar}$ in the Schrödinger equation (3.1) yields the coupled nonlinear partial differential equations (3.7) and (3.8) for R and S on $M\times I$. It follows from an elementary computation using these two equations that $\varphi(x,t)=e^{(R+S)(x,t)/\hbar}$ satisfies the dynamical equation (4.24) and then reduces indeed to $\varphi_{T/2}(x)$ at $t=T/2$. The analogous argument for the forward evolution shows that $\varphi^*(x,t)=e^{(R-S)(x,t)/\hbar}$ satisfies the parabolic equation (4.23) and reduces to the correct initial condition $\varphi^*_{-T/2}(x)$ at $t=-T/2$. By uniqueness for the solution of the Schrödinger system, the conclusion holds.

Let us summarize some useful information about the (real time) Markovian Bernstein diffusion process X_t.

Corollary 4.3.1: In the conditions of the Theorem 4.3, we have the following.

(1) The probability density of the (unique) Markovian Bernstein process X_t on I is given by (strictly positive) continuous function on $L^1(M,dx)$,

$$\rho(x,t)dx=e^{2R(x,t)}\,dx=|\psi(x,t)|^2\,dx.$$

(2) The densities of its forward and backward transition probabilities are

$$p(s,x,t,y)=k(s,x,t,y)\,\frac{e^{(R+S)(y,t)/\hbar}}{e^{(R+S)(x,s)/\hbar}}$$

and

$$p_*(s,x,t,y)=\frac{e^{(R-S)(x,s)/\hbar}}{e^{(R-S)(y,t)/\hbar}}k(s,x,t,y),$$

$$-T/2\leqslant s<t\leqslant T/2,\quad (x,y)\in M\times M.$$

(3) The forward and backward drifts, and the diffusion matrix of the process are, respectively,

$$b(x,t)=(\nabla R+\nabla S)(x,t),\quad x\in M,\quad t\in I,$$

$$b_*(x,t)=(-\nabla R+\nabla S)(x,t),$$

$C(x,t)=C_*(x,t)=\hbar I$, for I the $n\times n$ identity matrix.

Proof: (1) and (2) are immediate consequences of Corollary 3.3.1 for $\varphi^*(x,t)=e^{(R-S)(x,t)/\hbar}$ and $\varphi(x,t)=e^{(R+S)(x,t)/\hbar}$, and (3) of Corollary 3.3.2.

The comparison with the diffusion process described in the Introduction [(1.4) and (1.5)] shows that we have found indeed a new construction of the process of stochastic mechanics. Like the one in imaginary time, it is obviously valid for all times.

The description of the dynamical part of stochastic mechanics is analogous to the one given in Sec. IV A, so we go faster, and we assume that we are in the conditions of the Theorem 4.3.

Theorem 4.4: Let $b=\hbar\nabla\varphi/\varphi$ and $b_*=-\hbar\nabla\varphi^*/\varphi^*$ be the forward and backward drift of the Markovian Bernstein process X_t (Corollary 4.3.1). If we define the current velocity v and the osmotic velocity u of this process by the M valued functions on $M\times I$.

$$v=\hbar\nabla\log(\varphi/\varphi^*)^{1/2}\tag{4.26}$$

and

$$u=\hbar\nabla\log(\varphi^*\varphi)^{1/2}=\hbar\nabla\log\rho^{1/2}\tag{4.27}$$

the dynamics of the process is described by the coupled nonlinear partial differential equations on $M\times I$,

$$\begin{aligned}\frac{\partial u}{\partial t}&=-\frac{\hbar}{2}\operatorname{grad}\operatorname{div}v-\operatorname{grad}v\cdot u,\\ \frac{\partial v}{\partial t}&=-\frac{\hbar}{2}\Delta u-u\nabla u-v\nabla v+\nabla\vartheta,\end{aligned}\tag{4.28}$$

and the change of dependent variable $\varphi^*=e^{(R-S)/\hbar}$ transforms the system (4.28) into the parabolic equation (4.23).

Proof: Using φ and φ^* like θ and θ^* in Sec. IV A, and the probability density $\rho=\varphi^*\varphi$ like p, one verifies easily that the local conservation of probability means that ρ is a weak solution on $M \times I$ of

$$\frac{\partial \rho}{\partial t} = \operatorname{div} j = 0\,, \qquad (4.29)$$

where the current j is defined by

$$j = (\hbar/2)[\varphi^* \nabla \varphi - \varphi \nabla \varphi^*]\,. \qquad (4.30)$$

The argument is the one of Proposition 4.1. This justifies the definition of a current velocity v such that $j = \rho v$, therefore v reduces to (4.26). If the osmotic velocity u is defined by (4.27), the Corollary 4.3.1 (3) shows that

$$b = v + u = \nabla S + \nabla R \qquad (4.31)$$

and

$$b_* = v - u = \nabla S - \nabla R\,. \qquad (4.32)$$

It follows from the equations of motion of φ^* and φ on $M \times I$, (4.23) and (4.24), that the evolution of u and v is described by the system (4.28).

Since $\varphi^* = e^{(R-S)/\hbar}$ on $M \times I$, this is a linearization of the system (4.28). Actually, we already know that φ^* satisfies the parabolic equation (4.23) by construction. □

In real time, it is worthwhile to observe that the above-mentioned linearization of the system (4.28) is not the most interesting one from the physical point of view.

Corollary 4.4.1: In the same conditions, the change of dependent variable $\psi = e^{(R+iS)/\hbar}$ transforms the system (4.28) into the starting Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2} \Delta \psi + V\psi\,. \qquad (4.33)$$

Proof: It is sufficient to remark that the modified potential ϑ of the starting parabolic differential equation [(3.3) and (3.4)] may be written, using $u = \nabla R$ [Eqs. (4.31) or (4.32)], as

$$\vartheta = u^2 + \hbar \operatorname{div} u - V\,. \qquad (4.34)$$

Therefore the second equation of the system (4.28) is modified to

$$\frac{\partial v}{\partial t} = \frac{\hbar}{2} + \Delta u + \frac{1}{2} \operatorname{grad} u^2 - \frac{1}{2} \operatorname{grad} v^2 - \nabla V\,. \qquad (4.35)$$

This relation is the dynamical equation found by Nelson in his original derivation of stochastic mechanics.[3] We already know that the change of variable $\psi = e^{(R+iS)/\hbar}$ in the Schrödinger equation (4.33) gives the coupled nonlinear system [(3.7) and (3.8)]. Taking gradients, it follows from the relations (4.31) and (4.32) that this system reduces to the first equation of (4.28) and to (4.35). □

To conclude the analysis in real time, we have to show, as in Sec. IV A, why the Markovian Bernstein process X_t is naturally associated to a mechanics.

Corollary 4.4.2: In the same conditions as before and, in particular, assuming that V is of class C^1, Eq. (4.35) is equivalent to the Newton equation

$$\tfrac{1}{2}(DD_* X + D_* DX)(t) = -\nabla V\,. \qquad (4.36)$$

Proof: As in imaginary time, this is a straightforward computation. By (4.31) and (4.32),

$$DX = v + u$$

and

$$D_* X = v - u\,.$$

Using (2.6),

$$\begin{aligned} D_* DX &= D_* v + D_* u \\ &= \frac{\partial v}{\partial t} + (v-u)\nabla v - \frac{\hbar}{2}\nabla v + \frac{\partial u}{\partial t} \\ &\quad + (v-u)\nabla u - \frac{\hbar}{2}\Delta u\,, \end{aligned}$$

and, according to (2.5),

$$\begin{aligned} DD_* X &= Dv - Du \\ &= \frac{\partial v}{\partial t} + (v+u)\nabla v + \frac{\hbar}{2}\Delta v - \frac{\partial u}{\partial t} \\ &\quad - (V+u)\nabla u - \frac{\hbar}{2}\Delta u\,. \end{aligned}$$

Therefore

$$\frac{1}{2}(DD_* X + D_* DX) = \frac{\partial v}{\partial t} + v\nabla v - u\nabla u - \frac{\hbar}{2}\Delta u\,,$$

and Eq. (4.35) reduces to (4.36). □

Thus, in real time, we also obtain a version of stochastic mechanics for which the left-hand side of Eq. (4.36) defines a natural notion of acceleration. The sign of the right-hand side is correct in the sense that at the classical limit of differentiable trajectories the equation reduces to the classical Newton equation. However, the comparison with the Newton equation in imaginary time (4.36) shows that the notion of acceleration has to be modified when we go from imaginary to real time, in contrast, obviously, with the classical situation. This suggests that the analytical continuation in the time shall have much more interesting features for stochastic dynamics than for classical ones.[12,19]

Remarks: (1) In the construction proposed here the current and osmotic velocities $\overline{V}$ and $\overline{U}$ (or v and u) play a symmetrical role from the beginning. That is why, for example, the fact that the current velocity $\overline{V}$ (or v) is a gradient [Eqs. (4.8) and (4.26)] is no longer an assumption as in the original construction of Nelson[3] but a consequence of our constructive approach.

(2) The two notions of stochastic acceleration associated to the imaginary time and real time versions of stochastic mechanics [Eqs. (4.18) and (4.36)] are the only two time-symmetrical candidates for the title of mean acceleration, using the forward and backward derivatives (2.1) and (2.4). This point was observed by Nelson[2] but no dynamical meaning was assigned by him to the imaginary time possibility.

(3) The comparison between the two realizations starting from Eq. (3.5) (Sec. IV A) and Eq. (3.6) (Sec. IV B) is now quite easy. The first one is easier to understand dynamically: nothing else than the physical potential V is involved in the construction. But we cannot give an explicit description of the associated process Z_t before solving the (complicated) Schrödinger system (4.1).

In real time, we are able to avoid having to find the solution of Eq. (4.21) because we already use a *solution* of the starting equation (4.33). The price to pay is that the probabilistic characterization of the process X_t involves necessarily (Corollary 4.3.1) the modified potential ϑ, in spite of the validity of the Newton equation (4.36). This means that for a given physical potential V, the class of resulting processes X_t have few common probabilistic properties.

C. Mechanical stationary states

Suppose that the Hamiltonian of the Schrödinger equation (3.1) (for $\sigma = i$) satisfies the hypothesis of the Theorem 3.1 and has a ground state (this is a hypothesis on the potential V). Then it is well known that this ground state is unique and can be chosen to be strictly positive. Let us denote by $\rho_0 = \rho_0(x)$ the invariant density of the homogeneous diffusion process associated to this situation by real time stochastic mechanics.

This implies that the given solution of the Schrödinger equation takes the form

$$\psi_0(x,t) = \rho_0^{1/2}(x) e^{-iE_0 t/\hbar}, \tag{4.37}$$

with $\rho_0^{1/2}(x) \in L^2(M)$ (here $M = \mathbb{R}^n$) and E_0 is an isolated eigenvalue, the inferior bound for the spectrum of the Hamiltonian H. According to (3.2), this means that

$$R(x,t) = \hbar \log \rho_0^{1/2}(x)$$

and

$$S(x,t) = -E_0 t.$$

If we introduce the (smooth) function

$$b_0(x) = \nabla R,$$

the modified potential ϑ of the parabolic equation (3.3) is

$$\vartheta(x) = b_0^2 + h \operatorname{div} b_0 - V.$$

But the (time-independent) Schrödinger equation means precisely

$$V - E_0 = (\hbar/2) \operatorname{div} b_0 + \tfrac{1}{2} b_0^2,$$

and we find, after substitution of this in ϑ, that, up to an additive constant, the modified potential reduces to the physical one,

$$\vartheta = V. \tag{4.38}$$

Consequently, the strictly positive (since V is bounded below) fundamental solution k of the parabolic equation (3.3), in this stationary case, coincides with the fundamental solution h of the imaginary time construction [i.e., of Eq. (3.5)],

$$\begin{aligned} k(s,x,t,y) &= \text{kernel } T_{t-s} \\ &= h(s,x,t,y). \end{aligned} \tag{4.39}$$

It is immediate to verify that the (unique) solution of the associated Schrödinger system (3.26)≡(3.27) for the invariant boundary probability density ρ_0 reduces to

$$\varphi^*_{-T/2}(x) = \rho_0^{1/2}(x) e^{E_0 T/2\hbar}$$

and (4.40)

$$\varphi_{T/2}(y) = \rho_0^{1/2}(x) e^{-E_0 T/2\hbar},$$

and then, using Corollary 4.3.1, the forward probability of the diffusion process with drift b_0 to

$$p(s,x,t,y) = h(s,x,t,y) \frac{\rho_0^{1/2}(y) e^{-E_0 t/\hbar}}{\rho_0^{1/2}(x) e^{-E_0 s/\hbar}}. \tag{4.41}$$

As soon as we consider an excited state of the Schrödinger equation the construction seems to break down. In this case, the given invariant boundary probability densities of the Schrödinger system have nodes [because the associated wave function ψ_j in $L^2(M)$, $j>0$, has to change its sign in order to be orthogonal to ψ_0] and it was already observed by Fortet,[18] 40 years ago, that we lose the uniqueness of the solution for this system. Actually, this is perfectly natural from the physical point of view, and we proceed now to the construction of the unique Markovian Bernstein process corresponding to a stationary solution of the Schrödinger equation, smooth except at the nodes.

We consider only the one-dimensional case ($M = \mathbb{R}$) for the simplicity of exposition.

Let us assume that the potential V is such that the Schrödinger equation (3.1) ($\sigma = i$) admits a stationary solution of the form

$$\psi_j(x,t) = \phi_j(x) e^{-iE_j t/\hbar}, \tag{4.42}$$

for $\phi_j(x)$ a real element of $L^2(\mathbb{R})$ (it is always possible to do this) and E_j not the lowest eigenvalue of the Hamiltonian. Then $\phi_j(x) = \rho_j^{1/2}(x)$ changes its sign on the state space $M = \mathbb{R}$. (By convention, $\rho_j^{1/2}$ is the positive or negative square root, depending on the chosen region of the sate space.) We consider simultaneously two kinds of connected domains Λ of the state space (the real line), a semi-infinite interval $]z_1,\infty[\equiv \Lambda_1$ (or $]-\infty,z_n[$) for z_1 one of the two extreme zeros of $\phi_j(x)$, and a bounded interval $]z_1,z_2[\equiv \Lambda_{12}$ for z_i, $i = 1,2$, two consecutive zeros of $\phi_j(x)$. Actually, the standard one-point compactification of these sets (denoted in the same way) will be used. We shall construct a unique Markovian Bernstein diffusion process on each of these two kinds of domains, considered as state spaces of their own. In requiring natural (and compatible) boundary conditions of the border of two such domains, the process on all of the state space can be decomposed in this way.

To the given stationary solution of the Schrödinger equation $\psi_j(x,t)$, real time stochastic mechanics associates formally a homogeneous diffusion process with vanishing current velocity v (in the sense of Theorem 4.4) and drift

$$b_j(x) = \nabla \phi_j(x)/\phi_j(x), \tag{4.43}$$

which is clearly singular at the zeros of the wave function ψ_j. Let us denote by $\bar{\phi}_j$ the restriction of ϕ_j to a domain Λ, namely $\bar{\phi}_j(x) = \phi_j(x)\chi_\Lambda(x)$, where χ_Λ is the characteristic function of Λ.

It is easy to check that, as for the ground state, the modified potential ϑ for this excited state on Λ is nothing but (up to an additive constant) the restriction to Λ of the physical potential V. In other words, in any stationary situation, the two given versions of mechanics coincide and from now we can refer to stochastic mechanics, without further specification.

Let us consider the Kolmogorov backward (parabolic) equation for the homogeneous process on Λ,

$$\frac{\partial u}{\partial t} = b_j \nabla u + \frac{\hbar}{2} \Delta u \equiv A_j u, \tag{4.44}$$

where $b_j = \nabla\overline{\phi}_j/\overline{\phi}_j$ and A_j is the generator of the corresponding transition semigroup $P_t = e^{-tA_j}$. Notice that u in (4.44) has nothing to do with the osmotic velocity of Theorem 4.4: we simply conform here to the traditional probabilistic notation. Since the stochastic continuity of the transition probability was assumed, this semigroup is uniquely defined by A_j.

We choose to denote by $k_{\Lambda j}(s,x,t,y) = k_{\Lambda j}(x,t-s,y)$ the kernel of the starting Schrödinger semigroup $U_t \equiv T_t = e^{-tH}$ and by $P_{\Lambda j}(s,x,t,y) = P_{\Lambda j}(x,t-s,y)$ the integral kernel of P_t.

Before describing the relation between the two semigroups P_t and U_t, we have to characterize the domain of A_j, $\mathscr{D}(A_j)$, by the specification of physical boundary conditions for the process. If Λ is of the form $\Lambda_1 =]z_1,\infty[$, it is natural (since $\overline{\phi}_j^2$ is the quantum probability density in Λ_1) to require that

the set of the twice continuously differentiable functions u such that

$$A_j u \in C(\Lambda), \quad \overline{\phi}_j^2(x)\nabla u \underset{x\downarrow z_1}{\longrightarrow} 0, \quad A_j u(x) \underset{x\uparrow\infty}{\longrightarrow} 0$$

belong to $\mathscr{D}(A_j)$. (4.45)

If Λ is of the form $\Lambda_{12} =]z_1,z_2[$, we require that

the set of twice continuously differentiable functions u such that

$$A_j u \in C(\Lambda), \quad \overline{\phi}_j^2(x)u \underset{x\downarrow z_1}{\longrightarrow} 0, \quad \overline{\phi}_j^2(x)\nabla u \underset{x\uparrow z_2}{\longrightarrow} 0$$

belong to $\mathscr{D}(A_j)$. (4.46)

If $p_{\Lambda j}(x,t-s,y) = p_{\Lambda j}(y,t-s,x)$ and

$$\int_\Lambda\int_\Lambda p_{\Lambda j}^2(x,t-s,y)\,dx\,dy < \infty$$

then it is known that U_t has indeed an integral kernel $k_{\Lambda j}(x,t-s,y)$ s.t.

$$\int_\Lambda\int_\Lambda k_{\Lambda j}^2(x,t-s,y)\,dx\,dy < \infty .$$

Moreover, U_t is a positive self-adjoint operator of Hilbert–Schmidt type and there is a sequence $\{\mu_{j,m}\}_{m=1}^\infty$, $0 \leqslant \mu_{j,1} \leqslant \mu_{j,2} \leqslant \cdots$, with $\mu_{j,m} \to +\infty$, and an orthogonal basis of $L^2(\Lambda,dx)$, $\{\overline{a}_{j,m}\}_{m=1}^\infty$ such that

$$\overline{H}\,\overline{a}_{j,m} = \mu_{j,m}\,\overline{a}_{j,m} \tag{4.47}$$

and

$$U_t\,\overline{a}_{j,m} = e^{-\mu_{jm},t}\,\overline{a}_{j,m} . \tag{4.48}$$

Under these conditions, for fixed x, y in Λ, $t-s>0$,

$$k_{\Lambda j}(x,t-s,y) = \sum_m e^{-\mu_{j,m}(t-s)}\,\overline{a}_{jm}(x)\,\overline{a}_{j,m}(y)\pi_{j,m} , \tag{4.49}$$

where $\pi_{j,m} = [\int_\Lambda \overline{a}_{j,m}^2(x)dx]^{-1}$, is absolutely convergent.

Now $L^2(\Lambda,dx)$ is unitary equivalent to $L^2(\Lambda,\overline{\phi}_j^2(x)dx)$ under $U_j: g \mapsto g/\overline{\phi}_j$ and $\overline{H}$ to $(E_j - A_j)$. Therefore,

$$\overline{H} = U_j^{-1}(E_j - A_j)U_j$$

and $\{a_{j,m} \equiv \overline{a}_{j,m}/\overline{\phi}_j\}_{m=1}^\infty$ is an orthogonal basis of $L^2(\Lambda,\overline{\phi}_j^2(x)dx)$ such that

$$A_j a_{j,m} = -\lambda_{j,m} a_{j,m} \tag{4.50}$$

for the above-mentioned boundary conditions, and the relation between eigenvalues,

$$\mu_{j,m} = \lambda_{j,m} + E_j . \tag{4.51}$$

In other words the positive kernel k_j may also be expressed as

$$\begin{aligned} k_{\Lambda j}&(x,t-s,y)\\ &= \overline{\phi}_j(y)\overline{\phi}_j(x)e^{-E_j(t-s)}\\ &\quad\times \sum_{m=0}^\infty e^{-\lambda_{j,m}(t-s)}a_{j,m}(x)a_{j,m}(y)\pi_{j,m} . \end{aligned} \tag{4.52}$$

This is the starting kernel of our construction of Bernstein processes. According to Theorem 3.4, there is a unique solution for the associated Schrödinger system since we use the strictly positive boundary (invariant) probability densities on Λ,

$$\rho_j\Big(x,-\frac{T}{2}\Big) = \rho_j\Big(x,\frac{T}{2}\Big) = \frac{\overline{\phi}_j(x)}{\int_\Lambda \overline{\phi}_j^2(\xi)d\xi} . \tag{4.53}$$

It is easy to find this solution explicitly,

$$\begin{aligned} \varphi^*_{-T/2}(x) &= \overline{\phi}_j(x)e^{E_j(T/2)/\hbar} ,\\ \varphi_{T/2}(y) &= \overline{\phi}_j(y)e^{-E_j(T/2)/\hbar} , \end{aligned} \tag{4.54}$$

and therefore the forward transition probability density of the (unique) Bernstein process in Λ follows from Corollary 4.3.1,

$$\begin{aligned} p_{\Lambda j}&(x,t-s,y)\\ &= \overline{\phi}_j^2(y)\sum_{m=0}^\infty e^{-\lambda_{j,m}(t-s)}a_{j,m}(x)a_{j,m}(y)\pi_{j,m}, \end{aligned} \tag{4.55}$$

for x, y in Λ, $t-s>0$.

This is nothing but that the spectral representation for the density of the (forward) transition probability of the homogeneous diffusion process on Λ, with drift $b_j = \nabla\overline{\phi}_j/\overline{\phi}_j$. It is therefore possible to use the natural decomposition of the original state space M (here $M = \mathbb{R}$) in disjoint domains Λ, due to the nodes of the quantum boundary invariant probability density, for the construction of a unique stationary Bernstein process in each of these domains.

The two given sets of boundary conditions introduced in the eigenvalue problem of A are sufficient for this purpose. Using the classification, due to Feller, of the boundary conditions for the one-dimensional diffusion process,[20] it is easy to clarify the probabilistic sense of the two given conditions.

For a semi-infinite interval $\Lambda_1 =]z_1,\infty[$ the left boundary is an entrance boundary and then cannot be reached from the interior of the interval. In other words, the quantum node in z_1 is never reached. The right boundary is a natural boundary and therefore can neither be reached in finite expected time, nor may be a starting point of the process.

For a bounded interval of the form $\Lambda_{12} =]z_1,z_2[$, both boundaries are entrance.

This description agrees with the results known from the usual construction of the diffusion processes associated to the bound states.[1,21]

Notice that our result does not contradict the uniqueness of the process in the (analytical) sense of Carlen on all of the state space M (see Ref. 5). There is indeed only one process on M, but our description is more detailed and gives a different transition probability on each connected domain Λ between the nodes of the wave function.

The necessity to change the starting kernel of our construction for each region Λ is clearly due to the end-point effects of the nodes.

Remark: The method proposed here enables us to construct diffusion processes whose drifts are much more singular than the ones authorized by the general Theorems about existence and uniqueness of diffusion processes. An example is given in Ref. 12.

V. VARIATIONAL PROCESSES AND STOCHASTIC MECHANICS

As observed by Bernstein,[11] the hypothesis used for the construction of all these time-symmetric processes suggest that they may be accessible to a variational characterization, in analogy with the trajectories of classical mechanics, for two fixed end points. However, it was never done, probably because the natural dynamical meaning of these processes was not investigated.

Now, according to Proposition 4.2 and Corollary 4.4.2, this dynamical meaning for a specified potential V, is contained in two versions of Newton equations, in imaginary and real time.

As mentioned in the Introduction, we know already, thanks to Yasue's result[6,7] that the real time dynamics is indeed the result of a variational principle with an analog of our two fixed end points condition. In this section, we describe another variational approach associated to the structure of Bernstein processes itself, and therefore common to the imaginary and real time dynamics.

Let us consider the imaginary time dynamics, namely the Newton equation of Proposition 4.2.

In classical mechanics, the action functional associated with the Newton equation ($M=1$) by the Hamiltonian least action principle is

$$A_0[X] = \int_{-T/2}^{T/2} \left\{ \frac{1}{2} |\dot{X}(t)|^2 - V(X(t)) \right\} dt, \tag{5.1}$$

where $|\cdot|$ is the Euclidean norm.

Since the imaginary time dynamics is obtained from the real time one by an analytical continuation replacing t by $-it$ [cf. (3.1)], the action in imaginary time is proportional to

$$\overline{A}_0[z] = \int_{-T/2}^{T/2} \left\{ \frac{1}{2} |\dot{z}(t)|^2 + V(z(t)) \right\} dt. \tag{5.2}$$

If we denote as before the imaginary time Markovian Bernstein diffusion process by Z_t, $t \in I$, it is quite natural to consider the following stochastic generalization of the action (5.2):

$$\overline{A}_\hbar[Z] = E\left[\int_{-T/2}^{T/2} \left\{ \frac{1}{2} |D_* Z(t)|^2 + V(Z(t)) \right\} dt \right]. \tag{5.3}$$

Notice that the two actions coincide at the classical limit $\hbar = 0$. By analogy with the classical case, we will use the integrand (the "Lagrangian") of this action $\overline{A}_\hbar[Z]$ to construct a new action $\overline{A}_\hbar(x,t)$, a function of the future position x and time t [this is the reason for the choice of the backward derivative in (5.3)], intrinsically associated to the dynamics of Z_t. To do so, let us start from the (forward) parabolic equation used for the construction of Z_t, namely the Cauchy problem on $M \times I$,

$$-\hbar \frac{\partial \theta^*}{\partial t} = -\frac{\hbar^2}{2} \Delta \theta^* + V\theta^*, \tag{5.4}$$

where

$$\theta^*(x, -T/2) = \theta^*_{-T/2}(x) = e^{(\overline{R}_{-T/2} - \overline{S}_{-T/2})(x)/\hbar}$$

is the solution of the Schrödinger system (4.1) for some given boundary probabilities. In these conditions, the unique positive solution $\theta^*(x,t)$ of the Cauchy problem (5.4) contains the forward probabilistic information about the process Z_t. This partial information is time asymmetrical in nature.

Let us define the new function on $M \times I$:

$$\overline{A}_\hbar(x,t) = -\hbar \log \theta^*(x,t). \tag{5.5}$$

By construction, $A_\hbar$ is a solution of the nonlinear partial differential equation on $M \times I$,

$$\frac{\partial \overline{A}_\hbar}{\partial t} - \frac{\hbar}{2} \Delta \overline{A}_\hbar + \frac{1}{2} (\nabla \overline{A}_\hbar)^2 - V = 0, \tag{5.6}$$

with

$$\overline{A}_\hbar\left(x, -\frac{T}{2}\right) = (\overline{S}_{-T/2} - \overline{R}_{-T/2})(x).$$

Equation (5.6) is much more complicated than Eq. (5.4) but the existence of the solution of this new Cauchy problem is guaranteed by the one of the original problem (5.4). The point of the change of variable (5.5) is that (5.6) is a dynamic programming equation whose solution is the minimum value of some action functional. We shall use here an adaptation of a strategy developed by Fleming in the context of optimal stochastic control.[22]

Notice that (5.6) may also be written as

$$\frac{\partial \overline{A}_\hbar}{\partial t} - \frac{\hbar}{2} \Delta \overline{A}_\hbar - \hbar(x, \nabla \overline{A}_\hbar, t) = 0, \tag{5.7}$$

for $\hbar$: $M \times M \times I \to \mathbb{R}$ defined by

$$\hbar(x,p,t) = -\tfrac{1}{2} |p|^2 + V. \tag{5.8}$$

Lemma 5.1: Regarded as a function of p, $h = h(p)$ is a strictly concave function. It may also be expressed by

$$h(p) = \min_{B_*} [\mathscr{L}(\overline{B}_*) - p\overline{B}_*] = -\tfrac{1}{2}|p|^2 + V, \tag{5.9}$$

where $\mathscr{L}(\overline{B}_*) = \mathscr{L}(x, \overline{B}_*, t)$ is the strictly convex $M \times M \times I \to \mathbb{R}$ function dual to h in the sense that

$$\mathscr{L}(\overline{B}_*) = \max_p [h(p) + p\overline{B}_*] = \tfrac{1}{2} |B_*|^2 + V(x). \tag{5.10}$$

The unique $\overline{B}_*$ achieving the minimum in (5.9), denoted by B_*, is

$$B_* = -h'(p) = p. \tag{5.11}$$

The proof is elementary.

Coming back to Eq. (5.7) and the h function defined there, it follows from Lemma 5.1 that the $\overline{B}_* \equiv B_*$ achieving the minimum in Eq. (5.9) is

$$B_*(x,t) = -h'(x,\nabla\overline{A}_\hbar,t)$$
$$= \nabla\overline{A}_\hbar(x,t)\,. \tag{5.12}$$

Moreover, the $\mathscr{L}$ function defined in (5.10) is precisely the Lagrangian of our stochastic action (5.3). Given that (5.9) holds, Eq. (5.7) modifies to

$$\frac{\partial\overline{A}_\hbar}{\partial t} - \frac{\hbar}{2}\Delta\overline{A}_\hbar - \min_{\overline{B}_*}\left\{\frac{1}{2}|\overline{B}_*|^2 + V - \nabla\overline{A}_\hbar\,\overline{B}_*\right\} = 0\,. \tag{5.7'}$$

For the connection with the stochastic least action principle, we need to define a large class of $\mathscr{F}_\tau$ continuous semimartingales in the sense of Sec. II, whose $\overline{B}_*$ are the drifts, and the diffusion coefficient is fixed as before.

The collection of admissible processes for the variational principle is defined as follows.

If D_x denote the class of admissible drifts, we require that to each $\overline{B}_* \in D_x$ is associated a decreasing filtration $\mathscr{F}_\tau$, and a $\mathscr{F}_\tau$-Wiener process $w_*(\tau)$ such that $\overline{B}_*(\cdot)$ is $\mathscr{F}$ adapted. There is also a process $\overline{Z}(\tau) \equiv Z^{\overline{B}_*}(\tau)$ adapted to $\mathscr{F}_\tau$ such that

$$\overline{Z}(t) = x$$

and

$$d\overline{Z}(\tau) = \overline{B}^*(\tau)d\tau + \sqrt{\hbar}\,dw_*(\tau)\,, \quad -T/2 \leqslant \tau < t < T/2\,. \tag{5.13}$$

Finally, it is required that

$$\int_{-T/2}^{t} |\overline{B}_*(\tau)|d\tau < \infty \text{ a.s.} \tag{5.14}$$

As mentioned at the end of Sec. II, such a diffusion process $\overline{Z}(\tau)$ is generally not Markovian, but Itô's formula and then (2.5) and (2.6) are still valid.

In particular, if we evaluate Eq. (5.7') on any process $\overline{Z}(\tau)$ in this class we obtain

$$\frac{\partial\overline{A}_\hbar}{\partial t} + \overline{B}_*\nabla\overline{A}_\hbar - \frac{\hbar}{2}\Delta\overline{A}_\hbar - \frac{1}{2}|\overline{B}_*|^2 - V \leqslant 0 \tag{5.15}$$

or, by the definition (2.6) for the backward derivative,

$$D_*\overline{A}_\hbar(\overline{Z}(\tau),\tau) \leqslant \tfrac{1}{2}|\overline{B}_*(\tau)|^2 + V(\overline{Z}(\tau))\,. \tag{5.16}$$

Now we compute $E_{x,t}[\int_{-T/2}^{t}(\ )d\tau]$ on both sides of (5.16). Here $E_{x,t} = E$ is the expectation for such a $\overline{Z}(\tau)$ [i.e., for imposed final data $\overline{Z}(t) = x$].

It is a theorem of Nelson (Ref. 2, p. 96) that in these conditions, the left-hand expectation of (5.16) reduces to $\overline{A}_\hbar(x,t) - E_{x,t}[\overline{A}_h(\overline{Z}(-T/2), -T/2)]$. Taking into account the initial condition given in (5.6), we get

$$\overline{A}_\hbar(x,t) \leqslant E_{x,t}\int_{-T/2}^{t}\left\{\frac{1}{2}|\overline{B}_*(\tau)|^2 + V(\bar{z}(\tau))\right\}d\tau$$
$$+ E_{x,t}\left[(\overline{S}-\overline{R})\left(\overline{Z}\left(-\frac{T}{2}\right), -\frac{T}{2}\right)\right]. \tag{5.17}$$

We denote by $\overline{I}(\hbar,x;\overline{B}_*)$ the right-hand term of this inequality. Observe that the Lagrangian is the one of the action (5.3). We shall see that $\overline{A}_\hbar(x,t)$ is the minimal value of $\overline{I}(\hbar,x;\overline{B}_*)$ in the class of admissible processes. Indeed, in the terminology of optimal control theory, one says that the control $\overline{B}_*(\tau)$ is obtained from the feedback control law B_* for final data $\overline{Z}(t) = x$ when there is a B_* in D_x such that

$$\overline{B}_*(\tau) = B_*(\overline{Z}(\tau),\tau)\,. \tag{5.18}$$

If we choose for B_* the function defined in (5.12), we obtain a particular admissible process, denoted by $Z(\tau)$. Its drift is

$$B_*(Z(\tau),\tau) = \nabla\overline{A}_\hbar(Z(\tau),\tau)\,. \tag{5.19}$$

By construction, this drift is the one for which the minimum was achieved in Eq. (5.7), namely the one for which inequality (5.15) reduces to equality (5.6). Therefore

$$\overline{A}_\hbar(x,t) = \overline{I}(\hbar,x;B_*)\,. \tag{5.20}$$

Moreover, since the solution of the Cauchy problem (5.4) was found equal to

$$\theta^*(x,t) = e^{(\overline{R}-\overline{S})(x,t)/\hbar} \tag{5.21}$$

in the notations of Sec. IV A [Eq. (4.16)] the optimal drift (5.19) modifies to, taking into account the definition (5.5),

$$B_*(x,\tau) = (\nabla\overline{S} - \nabla\overline{R})(x,\tau)$$
$$= (\overline{V} - \overline{U})(x,\tau)\,. \tag{5.22}$$

This is the backward drift of the imaginary time version of stochastic mechanics [Sec. IV A, (4.13)].

Therefore, we have proved the following stochastic least action principle.

Theorem 5.1: Let $\theta^*(x,t)$ be the unique positive solution of the forward Cauchy problem (5.4) on $M \times I$, whose initial condition is a solution of the Schrödinger system (4.1) for some given boundary probabilities

$$\theta^*(x,-T/2) = e^{(\overline{R}_{-T/2}-\overline{S}_{-T/2})(x)/\hbar}\,.$$

If $\overline{A}_\hbar(x,t) = -\hbar\log\theta^*(x,t)$ and $\overline{B}_* \in D_x$, then

$$\overline{A}_\hbar(x,t) \leqslant \overline{I}(\hbar,x;\overline{B}_*)\,, \tag{5.23}$$

where the action $\overline{I}$ is defined by

$$\overline{I}(\hbar,x;\overline{B}_*)$$
$$= E_{x,t}\int_{-T/2}^{t}\left\{\frac{1}{2}|D_*\overline{Z}(\tau)|^2 + V(\overline{Z}(\tau))\right\}d\tau$$
$$+ E_{x,t}(\overline{S}-\overline{R})\left(Z\left(-\frac{T}{2}\right), -\frac{T}{2}\right).$$

Moreover, the (backward) drift of stochastic mechanics is a particular drift $\overline{B}_* \equiv B_*$ in D_x associated to an admissible Markovian process $Z(\tau)$ and such that

$$\overline{B}_*(\tau) = (\overline{V} - \overline{U})(Z(\tau),\tau)\,.$$

For this process $Z(\tau)$, the equality is achieved in (5.23), namely

$$\overline{A}_\hbar(x,t) = \inf_{\overline{B}_* \in D_x}\overline{I}(\hbar,x;\overline{B}_*)\,. \qquad \square$$

In the classical case, the main interest of the least action principle is that it gives a characterization of the physical dynamics. This is also true here.

Corollary 5.1.1: A Markovian process $Z(\tau)$, which is the minimum of the action $\overline{I}(\hbar,x;\overline{B}_*)$ for $\overline{B}_* \in D_x$, satisfies the Newton equation in imaginary time of Proposition 4.2,

$$\tfrac{1}{2}(DDZ + D_*D_*Z)(\tau) = \nabla V\,, \quad -T/2 \leqslant \tau < t\,, \tag{5.24}$$

with the "boundary conditions"

$$D_* Z(-T/2) = (\nabla \bar{S}_{-T/2} - \nabla \bar{R}_{-T/2})(Z(-T/2)), \quad Z(t) = x. \tag{5.25}$$

Proof: The proof of Theorem 5.1 shows that on this Markovian process $Z(\tau)$, Eq. (5.16) reduces to the equality

$$D_* \bar{A}_\hbar(Z(\tau),\tau) = \tfrac{1}{2}|B_*(Z(\tau),\tau)|^2 + V(Z(\tau)),$$

namely, using (5.19),

$$\frac{\partial \bar{A}_\hbar}{\partial \tau} + \frac{1}{2}|\nabla \bar{A}_\hbar|^2 - \frac{\hbar}{2}\Delta \bar{A}_\hbar = V.$$

Applying ∇ and interchanging the order of differentiation,

$$\frac{\partial}{\partial \tau}\nabla \bar{A}_\hbar + \nabla \bar{A}_\hbar \cdot \nabla(\nabla \bar{A}_\hbar) - \frac{\hbar}{2}\Delta \nabla \bar{A}_\hbar = \nabla V.$$

By (5.19) this is $D_* B_* = \nabla V$ or, since $B_* = \bar{V} - \bar{U}$ by (4.13),

$$D_* \bar{V} - D_* \bar{U} = \nabla V.$$

According to the definition of D_*, this is

$$\frac{\partial \bar{V}}{\partial \tau} + (\bar{V} - \bar{U})\nabla \bar{V} - \frac{\hbar}{2}\Delta \bar{V} - \frac{\partial \bar{U}}{\partial t} - (\bar{V} - \bar{U})\nabla \bar{U} + \frac{\hbar}{2}\Delta \bar{U} = \nabla V.$$

But we know already by (4.15) that

$$\frac{\partial \bar{U}}{\partial \tau} = -\frac{\hbar}{2}\operatorname{grad}\operatorname{div}\bar{V} - \operatorname{grad}\bar{V}\cdot\bar{U}$$

and therefore

$$\frac{\partial \bar{V}}{\partial \tau} = -\frac{\hbar}{2}\Delta \bar{U} - \bar{U}\nabla \bar{U} - \bar{V}\nabla \bar{V} + \nabla V.$$

This is precisely the Newton equation (5.24), by Proposition 4.2. The given boundary conditions are consequences of the construction proposed in Theorem 5.1. □

Theorem 5.1 and its corollary confirm the results of Sec. IV. To the unique Markovian Bernstein process in imaginary time associated to a couple of given probabilities corresponds a dynamical equation, the Newton equation (5.24), and this one follows from a least action principle involving only the classical Lagrangian of the problem [cf. (5.3)]. This justifies, for this new theory of classical diffusing particles, the name of "(Schrödinger's) stochastic variational dynamics."

Remark: On the basis of Theorem 5.1, one expects the following kind of theorem for the "classical limit": If V is bounded below and C^1, then

$$\lim_{\hbar \to 0} -\hbar \log e^{-(t+T/2)H/\hbar} e^{-\bar{S}(x)_{-T/2}/\hbar} = \min_{X \in \Gamma} \int_{-T/2}^{t} \left\{\frac{1}{2}|\dot{X}(\tau)|^2 + V(X(\tau))\right\} d\tau + \bar{S}_{-T/2}\left(X\left(-\frac{T}{2}\right)\right),$$

where Γ is the set of C^1 paths with $X(t) = x$, $\dot{X}(-T/2) = \nabla \bar{S}_{-T/2}(X(-T/2))$. These kinds of results are known (see Ref. 14, VI 18) and Theorem 5.1 can be considered as the generalization of them.

We do not go further in this direction because it will be more natural to discuss briefly the classical limit in the quantum mechanical context. The next Theorem is the real time version of Theorem 5.1.

Theorem 5.2: Let $\varphi^*(x,t)$ the unique positive solution of the forward Cauchy problem (4.23) on $M \times I$,

$$-\hbar\frac{\partial \varphi^*}{\partial t} = -\frac{\hbar^2}{2}\Delta \varphi^* + \vartheta \varphi^*,$$

whose initial condition is the solution of the Schrödinger system (4.21)

$$\varphi^*(x, -T/2) = e^{(R_{-T/2} - S_{-T/2})(x)/\hbar},$$

for two given quantum boundary densities of probability [(4.19) and (4.20)].

If $\bar{A}_\hbar(x,t) = -\hbar \log \varphi^*(x,t)$ and $\not{b}_* \in D_x$, then

$$\bar{A}_\hbar(x,t) \leqslant I(\hbar,x; \not{b}_*), \tag{5.26}$$

where the action I is defined by

$$I(\hbar,x;\not{b}_*) = E_{x,t}\int_{-T/2}^{t}\left\{\frac{1}{2}|D_*\bar{X}(\tau)|^2 + \vartheta(\bar{X}(\tau),\tau)\right\}d\tau + E_{x,t}(S - R)\left(\bar{X}\left(-\frac{T}{2}\right), -\frac{T}{2}\right).$$

Moreover, the (backward) drift of real time stochastic mechanics is a particular drift $\not{b}_* = b_*$ in D_x associated to an admissible Markovian process $X(\tau)$ such that

$$\not{b}_*(\tau) = (v - u)(X(\tau),\tau).$$

For this process $X(\tau)$, the equality is achieved in (26), then

$$A_\hbar(x,t) = \inf_{\not{b}_* \in D_x} I(\hbar,x;\not{b}_*).$$

The proof is the same as the one of Theorem 5.1, for φ^* replacing θ^* and $\vartheta = \vartheta(x,t)$ replacing $V = V(x)$. □

Corollary 5.2.1: A Markovian process $X(\tau)$, which is the minimum of the action $I(\hbar,x;\not{b}_*)$ for $\not{b}_* \in D_x$, satisfies the Newton equation in real time of Corollary 4.4.2,

$$\tfrac{1}{2}(DD_*X + D_*DX)(\tau) = -\nabla V, \quad -T/2 \leqslant \tau < t, \tag{5.27}$$

with the "boundary conditions"

$$D_*X(-T/2) = (\nabla S_{-T/2} - \nabla R_{-T/2})(X(-T/2)), \quad X(t) = x. \tag{5.28}$$

Proof: For the Markovian process $X(\tau)$ we have

$$D_* A_h(X(\tau),\tau) = \tfrac{1}{2}|b_*(X(\tau),\tau)|^2 + \vartheta(X(\tau),\tau)$$

and, since $b_*(X(\tau),\tau) = \nabla A_h(X(\tau),\tau)$,

$$\frac{\partial A_\hbar}{\partial t} + \frac{1}{2}|\nabla A_\hbar|^2 - \frac{h}{2}\Delta A_h = \vartheta.$$

Applying ∇ and interchanging the order of differentiation,

$$\frac{\partial}{\partial \tau}\nabla A_\hbar + \nabla A_\hbar \cdot \nabla(\nabla A_\hbar) - \frac{\hbar}{2}\Delta \nabla A_\hbar = \nabla \vartheta,$$

namely $D_* b_* = \nabla \vartheta$ or, since $b_* = v - u$ [(4.32)], and $\vartheta = u^2 + \hbar \operatorname{div} u - V$ [(4.34)],

$$D_* v - D_* u = 2u\nabla u + \hbar \Delta u - \nabla V.$$

Using the definition of the backward derivative and Eq. (4.28) for $\partial u/\partial t$, we get

$$\frac{\partial v}{\partial t}=\frac{\hbar}{2}\,\Delta u+\frac{1}{2}\,\text{grad}\,u^2-\frac{1}{2}\,\text{grad}\,v^2-\nabla V.$$

This is the Newton equation (5.27) according to Corollary 4.4.2. As before, the boundary conditions (5.28) follow from the construction of Theorem 5.2. □

Let us consider the (absolute) expectation of the action found in Theorem 5.2 (without the boundary term, and on all of I),

$$A_\hbar[X]=E\left[\int_{-T/2}^{T/2}\left\{\frac{1}{2}\,|D_*X(t)|^2+\vartheta(X(t),t)\right\}dt\right]. \tag{5.29}$$

Notice that, since $\vartheta=u^2+\hbar\,\text{div}\,u-V$,

$$\int_M\vartheta\cdot\rho\,dx=\int_M u^2\rho\,dx+\hbar\int_M\text{div}\,u\cdot dx-\int_M V\rho\,dx$$
$$=-\int_M(u^2+V)\rho\,dx\,,$$

after integration by parts of the divergence term, and the use of $u=\hbar\nabla\log\rho^{1/2}$ [(4.27)]. Therefore

$$A_\hbar[X]=E\left[\int_{-T/2}^{T/2}\left\{\frac{1}{2}\,|D_*X|^2-u^2-V\right\}dt\right]. \tag{5.30}$$

The interesting point here, in contrast with the imaginary time situation Eq. (5.3), is that the relevant action for quantum mechanics [(5.30)] involves not only the classical Lagrangian, $\frac{1}{2}|\dot{X}|^2-V$, but a supplementary term $-u^2$, depending on the osmotic velocity. This difference with the imaginary time case has some interesting implications.[12] No definition of stochastic action relevant for a variational principle in stochastic mechanics and involving *exclusively* the classical Lagrangian is known today.

Remarks: (1) It may seem contradictory that the construction of the process X_t, $t\in I$, involves in a crucial way the modified potential ϑ and its associated integral kernel k [Corollary 4.3.1 and Eq. (5.29)], in contrast to the dynamical equation (5.27), in which only the physical potential V appears. There is no contradiction however. An easy computation shows that the expectations of the "force" and of the "torque" due to the nondynamical terms of ϑ are zero for any solution of the Schrödinger equation. So, these supplementary terms have no dynamical consequences.

(2) The action $I(\hbar,x;b_*)$ (without boundary conditions) is not the one proposed by Guerra and Morato[23] but enables us to obtain, under weaker hypotheses (non-Markovian admissible processes) stronger conclusions (minimum of the action and not only critical point) for the same dynamical content (Corollary 5.2.1).

(3) At the classical limit $\hbar=0$, the osmotic velocity u vanishes since the forward and backward drift b and b_* reduce to the classical velocity [cf. (4.31) and (4.32)]. Consequently the action (5.30) reduces to the classical one

$$A_0[X]=\int_{-T/2}^{T/2}\left[\frac{1}{2}\,|\dot{X}(t)|^2-V(X(t))\right]dt. \tag{5.31}$$

In the same limit the function $A_0(x,t)$ of Theorem (5.2) reduces to $S=S(x,t)$ and the nonlinear partial differential equation satisfied by S [the real time analog of (5.6)] is the Hamilton–Jacobi equation on $M\times I$

$$\frac{\partial S}{\partial t}=\frac{1}{2}(\nabla S)^2+V=0\,. \tag{5.32}$$

This one is the result of the (classical) minimization of

$$I(0,x)=\int_{-T/2}^{t}\left\{\frac{1}{2}\,|\dot{X}|^2-V\right\}d\tau+S_{-T/2}\left(X\left(-\frac{T}{2}\right)\right),$$
$$-T/2\leqslant t\leqslant T/2\,. \tag{5.33}$$

Of course, the solution of this classical variational problem is equivalently characterized as the C^2 path such that $\ddot{X}=-\nabla V$, with boundary conditions

$$\dot{X}\left(-\frac{T}{2}\right)=\nabla S_{-T/2}\left(X\left(-\frac{T}{2}\right)\right)\quad\text{and}\quad X(t)=x\,. \tag{5.34}$$

This is indeed the "classical limit" of Corollary 5.2.1.

(4) Since in Theorems 5.1 and 5.2 the class of admissible processes includes, in particular, non-Markovian processes, the results of this section mean that no non-Markovian continuous semimartingale is dynamically relevant for any version of stochastic mechanics. In real time, this answers negatively to a conjecture of Nelson (Ref. 1, §23).

We are now ready to define informally the class of stochastic processes which motivates the title of this paper: A variational process on M, indexed by I, is the unique Markovian Bernstein process constructed according to Secs. III and IV, for two arbitrarily given boundary probabilities in imaginary time, and for a given pair of quantum boundary probabilities associated to a solution of Schrödinger equation, in real time. The dynamics on I of such a process, the Newton equations in imaginary time and real time, is characterized in a variational way by the least action principles of Sec. V. Therefore the variational processes can be interpreted as the natural generalizations of the trajectories of classical mechanics, considered from the variational point of view.

VI. CONCLUSION

In this paper, we are able to give a sense to the intuition of Schrödinger and Bernstein (cf. Appendix) according to which Bernstein processes are the stochastic analog of the classical trajectories from the variational point of view. This analogy is surprisingly complete since there are indeed stochastic generalizations of the Newton equation behind this construction and since, in real time, this generalization is indeed associated, according to the initial motivation of Schrödinger, to quantum dynamics. This provides a new constructive approach to stochastic mechanics, completely different from the one proposed initially by Nelson and realized recently by Carlen. It shows, in particular, that if a stochastic dynamics is determined by the data of an acceleration and of the physical potential, then the dynamical structure of stochastic mechanics is in no way restricted to the real time approach.

Actually, in spite of the dichotomic presentation adopted here (imaginary time versus real time) the construction directly inspired by Schrödinger ["(Schrödinger's) stochastic variational dynamics"] is already sufficient for the description of quantum phenomena. Since the proof involves some new technicalities, this will be shown elsewhere.[17]

From the mathematical point of view, the most interesting problem left open by this new constructive approach is to verify that it is valid under very general regularity conditions.

From the physical point of view, it will be interesting to investigate systematically the common points and the differences between the real time and the imaginary time versions of stochastic mechanics. We are convinced that a number of questions of interpretation, in stochastic mechanics, and in quantum mechanics, will be clarified in using this comparative point of view. Since the true stake of the discussion about the real time and imaginary time approaches of quantum phenomena lies in quantum field theory where the Euclidean point of view has been, until now, mathematically if not physically successful,[15] the results of such a comparative analysis will also have some interesting consequences at this level.

Finally, the fact that this constructive approach of quantum dynamics is the only one involving the data of two probabilities is not without consequences on the physical interpretation of the theory. Some of the elements relevant from this point of view are presented in Ref. 12.

ACKNOWLEDGMENTS

The author would like to thank Professor E. Nelson for his warm hospitality in Princeton and for his encouragements during the gestation of this work. It was also a pleasure to benefit from the comments of Professor R. Carmona, Dr. J. Lafferty, Professor P. A. Meyer, Professor P. Ruggiero, and Professor K. Yasue.

This research was supported in part by the Swiss National Science Foundation.

APPENDIX: BERNSTEIN PROCESSES

Fifty years ago, in two unnoticed papers,[10,17] Schrödinger described a very unusual way to look at diffusion processes. His motivation was to find an interpretation of his wave equation in classical probabilistic terms. He did not realize finally this program but got at least, according to his own words: "une analogie avec la mécanique ondulatoire qui fut si frappante...qu'il m'est difficile de la croire purement accidentelle." [10] The radically new aspect of the Schrödinger approach is to consider a diffusion process on a finite time interval $I = [-T/2, T/2]$, for a given initial probability density $\rho_{-T/2}(x)$ *and* a given final density $\rho_{T/2}(x)$ generally distinct from the one corresponding to the evolution of $\rho_{-T/2}(x)$ by the associated diffusion equation. His surprising conclusion was that the most probable evolution between these two boundaries is given by a time-reversible density of probability.

A few times later, Bernstein[11] gave a probabilistic interpretation of Schrödinger's idea and proposed a very specific program of construction of the involved new class of stochastic processes. In order to preserve the time reversibility of their evolution, he substituted to the usual construction of Markov processes for one in which (1) the initial density is replaced by the joint density m of the initial and final positions during the time interval I; (2) the transition of probability is replaced by a probability of passage K involving three positions, whose meaning is that one needs an information on the past $\mathscr{P}_t$ and one on the future $\mathscr{F}_t$ of the process to predict its dynamics in between; and (3) the Markov property is replaced by the condition that the knowledge of the position before a fixed past position X_t and after a fixed future position X_v does not modify the expectation (the probability) of the intermediate dynamics.

Bernstein called "reciprocal processes" the result of this construction. The program of Bernstein was realized by Jamison,[15] with crucial contributions of Beurling,[24] and Fortet.[18] We summarize the basic properties of these Bernstein processes following Jamison. We do not use the adjective "reciprocal" hereafter because we prefer to introduce the term "variational" to denote the relevant Bernstein processes for stochastic mechanics (Sec. V).

Let $X_t: \underset{\omega \mapsto \omega(t) = X(t,\omega)}{\Omega \to \dot{M}}$ be an M-valued stochastic process defined, according to Sec. II, on (Ω, σ_I, P).

Two subsigma algebras of σ_I are particularly important: the past at time t, $\mathscr{P}_t = \sigma\{X_s,\ s \leqslant t\}$, and the future at time t, $\mathscr{F}_t = \sigma\{X_u,\ u \geqslant t\}$. The present at time t, denoted by $\mathscr{N}_t$ ($\mathscr{N}$ for "Now") is given by $\mathscr{N}_t = \mathscr{P}_t \cap \mathscr{F}_t$. For two fixed times t and $v > t$ in the finite time interval $\mathscr{I}$, we introduce two other filtrations, $\mathscr{E}_{t,v} = \mathscr{P}_t \cup \mathscr{F}_v$ ($\mathscr{E}$ for "exterior") and $\mathscr{I}_{t,v} = \mathscr{F}_t \cap \mathscr{P}_v$ ($\mathscr{I}$ for "interior"). In particular, $\mathscr{I}_{t,t} = \mathscr{N}_t$ and $\mathscr{E}_{t,t} = \sigma_I$, the Borel sigma algebra of Ω.

Definition: X_t has the Bernstein property, if, for any $-T/2 < t < v < T/2$, for any A adapted to $\mathscr{E}_{t,v}$, and B adapted to $\mathscr{I}_{t,v}$,

$$P(A \cap B \mid X_t, X_v) = P(A \mid X_t, X_v) \cdot P(B \mid X_t, X_v)\,. \tag{A1}$$

As in the case of Markovian processes, some equivalent definitions are more useful. In particular, for any bounded Borel measurable g, and $-T/2 < t < u < v < T/2$,

$$E[g(X_u) \mid \mathscr{E}_{t,v}] = E[g(X_u) \mid X_t, X_v]\,. \tag{A2}$$

Here (A2) is clearly the above-mentioned condition (3) of Bernstein.

Lemma A.1: If X_t, $t \in I$, is Markovian, then it is a Bernstein process.

The proof can be found in Ref. 15.

Notice that the reciprocal assertion is not true.

Now we sketch the Jamison's proof of Theorem 3.2 for the existence and uniqueness of Bernstein process.

Proof of Theorem 3.2: (a) Uniqueness: Suppose that there is a probability measure P for the Bernstein process X_t such that (1) and (2) are true. For $-T/2 \leqslant t_1 < \cdots < t_n = T/2$, $B_i \in \mathscr{B}$, one verifies by induction on n that

$$\begin{aligned} &P(X_{t_2} \in B_2, \ldots, X_{t_n} \in B_n \mid X_{t_1}, X_{T/2}) \\ &\quad = \int_{B_2} K\left(t_1, X_{t_1}; t_2, dx_2; \frac{T}{2}, X_{T/2}\right) \int_{B_3} K\left(t_2, X_2; t_3, dx_3; \frac{T}{2}, X_{T/2}\right) \\ &\qquad \cdots \int_{B_{n-1}} K\left(t_{n-2}, X_{n-2}; t_{n-1}; dx_{n-1}; \frac{T}{2}, X_{T/2}\right) \int_{B_n} K\left(t_{n-1}, X_{n-1}; t_n, dx_n; \frac{T}{2}, X_{T/2}\right). \end{aligned} \tag{$*$}$$

This expression is denoted by $h(X_{t_1}, X_{T/2})$. Indeed for $n=2$,

$$P(X_{t_2}\in B_2|X_{t_1}, X_{T/2}) = K(t_1, X_{t_1};t_2,B_2;T/2,X_{T/2})$$

is true by hypothesis (2). Now suppose (*) is true for $n-1$, then

$$P(X_{t_2}\in B_2,...,X_{t_n}\in B_n|X_{t_1}, X_{T/2}) \equiv E\,[\chi_{B_2}(X_{t_2})P(X_{t_3}\in B_3,...,X_{t_n}\in B_n|X_{t_1}, X_{t_2}, X_{T/2})|X_{t_1}, X_{T/2}]\,.$$

With the Bernstein property (A2) and the above-mentioned notation for (*),

$$= E\,[\chi_{B_2}(X_{t_2})h(X_{t_2}, X_{T/2})|X_{t_1}, X_{T/2}]\,,$$

namely, by (2),

$$= \int_{B_2} K\left(t_1, X_{t_1};t_2,dx_2;\frac{T}{2}, X_{T/2}\right)h(X_{t_2}, X_{T/2})\,,$$

or, by the induction hypothesis,

$$= \int_{B_2} K\left(t_1, X_{t_1};t_2,dx_2;\frac{T}{2}, X_{T/2}\right)\int_{B_3} K\left(t_2, X_{t_2};t_3,dx_3;\frac{T}{2}, X_{T/2}\right)\cdots\int_{B_n} K\left(t_{n-1}, X_{t_{n-1}};t_n,dx_n;\frac{T}{2}, X_{T/2}\right).$$

This is indeed (*). Now, using this result for $t_1 = -T/2$ (and $t_2 = t_1$) in

$$P(C) = E\,[\chi_{B_s}(X_{-T/2})P(X_{t_1}\in B_1,...,X_{t_n}\in B_n|X_{-T/2}, X_{T/2})\chi_{B_F}(X_{T/2})]\,,$$

we get (3) and therefore the uniqueness of P_m.

(b) Existence of P_m: Let $\Omega_{[-[} \equiv \{\omega: I[-[\to M\}$ and $\sigma[-[\equiv \sigma\{X_t, t\in I[-[$, where $I[-[= [-T/2,T/2[$, and the probability space $(M\times M, \mathscr{B}\times\mathscr{B},m)$ for the starting and ending random variables. On this space $X_{-T/2}(x,y) = x$ and $X_{T/2}(x,y) = y$ are two random variables. If one denotes by $P_{-T/2}(B_s,y)$ the conditional probability $m(X_{-T/2}\in B_s|X_{T/2} = y)$ for B_s in $\mathscr{B}$ and each fixed z in M, $P_y(s,x,t,B) \equiv K(s,x;t,B;T/2,y)$ is a (forward) Markov transition probability. Indeed by (K3) for $A = M$, and (K1), the Chapman–Kolmogorov equation holds:

$$P_y(s,x,u,B) = \int_M P_y(s,x,t,d\xi)P_y(t,\xi,u,B)\,.$$

With $P_{-T/2}(\cdot, y)$ as initial distribution, the theory of Markov processes and the Kolmogorov's fundamental theorem assert the existence of a measure $\overline{P}_y$ on $\sigma_{[-[}$ such that with respect to $(\Omega_{[-[},\sigma_{[-[}, \overline{P}_y)$, X_t is the Markov process with these initial distribution and transition.

For $C_{[-[}$ a cylinder event, $\{X_{-T/2}\in B_s,...,X_{t_n}\in B_n\}\in\sigma_{[-[}$ and

$$\int_M\int_{B_E} m(x,y)dx\,dy \equiv P_{T/2}(X_{T/2}\in B_E)$$

the distribution of $X_{T/2}$, define on $(\sigma_{[-[}\times\mathscr{B})$ the probability

$$P_m(C_{[-[}\times B_E) = \int_{B_E}\overline{P}_y(C_{[-[})dP_{T/2}(y)\,. \qquad \text{(A3)}$$

It is the good one, after the identification of (Ω,σ_I) with the product space $(\Omega_{[-[}\times M,\sigma_{[-[}\times\mathscr{B})$.

(c) Validity of (3) and (1): Given the usual Markovian finite-dimensional distribution

$$\overline{P}_y(C_{[-[}) = \int_{B_S} m(X_{-T/2}\in dx|X_{T/2} = y)\int_{B_1} P_y\left(-\frac{T}{2},x,t_1,dx_1\right)\cdots\int_{B_n} P_y(t_{n-1},dx_{n-1},t_n,dx_n)$$

for (A3), we get

$$P_m(C) = \int_{B_E}\int_{B_S} m(X_{-T/2}\in dx|X_{T/2} = y)\int_{B_1} P_y\left(-\frac{T}{2},x,t_1,dx_1\right)\cdots\int_{B_n} P_y(t_{n-1},dx_{n-1},t_n,dx_n)dP_{T/2}(y)\,,$$

but this is (3) since, if $c(x,y)$ is the bounded $\mathscr{B}\times\mathscr{B}$ measurable function associated to the "interior cylinder,"

$$c(x,y) = \int_{B_1} P_y\left(-\frac{T}{2},x,t_1,dx_1\right)\cdots\int_{B_n} P_y(t_{n-1},x_{n-1},t_n,dx_n)\,,$$

we have the identity

$$\int_{B_S\times B_E} c(x,y)dm(x,y) \equiv \int_{B_E}\int_{B_S} c(x,y)m(X_{-T/2}\in dx|X_{T/2} = y)dP_{T/2}(y)\,.$$

Notice also that (3) implies (1) trivially.

(d) Using (3) and (K3) one verifies that $P_m(X_u\in B\,|\mathscr{G}_{t,v}) = K(t,X_t;u,B;v,X_v)$ namely that X_t has the Bernstein property with respect to (Ω,σ_t, P_m).

Remarks: (1) The construction of P_m given in (b) is also valid if we start from $\Omega_{]-]} \equiv \{\omega: I_{]-]} \to M\}$, $\sigma_{]-]}$, and

$m(X_{T/2} \in B_E | X_{-T/2} = x) \equiv P_{T/2}(B_E, x)$. For each fixed x in M, $P_{*x}(t,B,u,z) \equiv K(-T/2,x;t,B;u,z)$ is also a Markov transition [put $B=M$ in (K3), then

$$P_{*_x}(t,A,v,z) = \int P_{*x}(t,A,u,y) P_{*x}(u,dy,v,z)] .$$

With $P_{T/2}(\cdot,x)$ as final distribution, construct the measure $\overline{P}_{*x}$ such that with respect to $(\Omega_{[-[}, \sigma_{[-[} \overline{P}_{*x})$, X_t is the corresponding Markov process. For

$$\int_M \int_{B_S} m(x,y) dx\, dy \equiv P_{-T/2}(X_{-T/2} \in B_S) ,$$

the distribution of $X_{-T/2}$, we get the backward version of (A3),

$$P_m(B_S \times C_{]-]}) = \int_{B_S} \overline{P}_{*x}(c_{]-]}) dP_{-T/2}(x) . \tag{A3$_*$}$$

The usual backward Markovian finite-dimensional distribution

$$\overline{P}_{*x}(C_{]-]}) = \int_{B_1} P_{*x}(t_1,dx_1,t_2,x_2) \cdots \int_{B_n} P_{*x}\left(t_n,dx_n,\frac{T}{2},y\right) \int_{B_E} m(X_{T/2} \in dy | X_{-T/2} = x)$$

is used in (A3$_*$) to obtain $(3)_*$.

(2) This construction, and in particular (A3) and $(\text{A3})_*$, shows that if $X_{-T/2}$ or $X_{T/2}$ has a probability measure concentrated at one point the Bernstein process is also Markovian. Actually, the class of Bernstein processes is strictly larger than the Markov class.

[1]E. Nelson, *Quantum Fluctuations* (Princeton U.P., Princeton, NJ, 1985).
[2]E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U.P., Princeton, NJ, 1967).
[3]E. Nelson, Phys. Rev. **150**, 1079 (1966).
[4]K. Ito, "Stochastic calculus," in *Springer Lecture Notes in Physics*, Vol. 39 (Springer, New York, 1975), p. 218.
[5]E. Carlen, Commun. Math. Phys. **94**, 293 (1984).
[6]K. Yasue, J. Math. Phys. **22**, 1010 (1981).
[7]K. Yasue, J. Funct. Anal. **41**, 327 (1981).
[8]W. A. Zheng and P. A. Meyer, "Quelques résultats de Mécanique stochastique," in *Séminaire de Probabilité XVIII* (Springer, Berlin, 1984).
[9]J. C. Zambrini, Int. J. Theor. Phys. **24**, 277 (1985).
[10]E. Schrödinger, Ann. Inst. H. Poincaré **2**, 269 (1932).
[11]S. Bernstein, "Sur les liaisons entre les grandeurs aléatoires," Verh. Int. Math. Zürich, Band 1 (1932).
[12]J. C. Zambrini, Phys. Rev. A **33**, 1532 (1986).
[13]M. Kac, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (Univ. California P., Berkeley, 1951), p. 189.
[14]B. Simon, *Functional Integration and Quantum Physics* (Academic, New York, 1979).
[15]B. Jamison, Z. Wahrsch. Gebiete **30**, 65 (1974).
[16]E. B. Dynkin, "Markov processes," Vols. I and II of *Grundlehren der Math. Wissensch.* (Springer, Berlin, 1965).
[17]E. Schrödinger, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl. **1931**, 144.
[18]R. Fortet, J. Math. Pures Appl. **IX**, 83 (1940).
[19]S. Albeverio, K. Yasue, and J. C. Zambrini, in preparation.
[20]W. Feller, Ann. Math. **55**, 468 (1952).
[21]S. Albeverio and R. Høegh-Krohn, J. Math. Phys. **15**, 1745 (1974).
[22]W. H. Fleming, Appl. Math. Optim. **4**, 329 (1978).
[23]F. Guerra and L. Morato, Phys. Rev. D **27**, 1774 (1983).
[24]A. Beurling, Ann. Math. **72**, 189 (1960).

Large-N solution of the Klein–Gordon equation

Ashok Chatterjee
Department of Theoretical Physics, Indian Association for the Cultivation of Science, Jadavpur, Calcutta 700032, India

(Received 29 January 1986; accepted for publication 9 April 1986)

An iterative $1/N$ expansion scheme is developed to solve the Klein–Gordon equation to obtain the energy spectrum of a scalar particle in a spherically symmetric potential. For the Coulomb potential, this approach is shown to yield the exact results.

I. INTRODUCTION

The method of large-N expansion has emerged in recent years as a very useful approximation scheme in nonrelativistic quantum mechanics[1-3] and the multicomponent model problems in quantum field theory[4,5] and statistical physics.[6,7] The leading-order approximation in this theory consists of assuming the number of degrees of freedom of the system to be infinitely large ($N\to\infty$) and leads in a number of cases to impressive results. However, the leading-order solution may not be sufficient in general[8] and one has to incorporate the finite-N corrections by introducing a systematic expansion in powers of $1/N$. In nonrelativistic quantum mechanics the method of large-N expansion serves as an alternative to ordinary perturbation theory and employs for spherically symmetric potentials $1/k = 1/(N+2l)$ as the expansion parameter, where N is the spatial dimensionality and l the angular momentum. Mlodinow and Papanicolaou[1] have shown by a group theoretic formulation that $1/k$ is indeed a natural expansion parameter. Thus a $1/N$ expansion is essentially a nonperturbative expansion and contains in it the promise of solving the strong coupling problems for which the usual perturbative treatments fail. In fact, in QCD the only meaningful expansion parameter is $1/N$, where N is the number of colors.

Recently, Miramontes and Pajares[9] have studied the large-N limits in relativistic quantum mechanics and have shown that for a class of potentials $V(r)\sim r^n$, where $-2<n<0$, the relativistic and spin corrections are nonleading in $1/N$ expansion. As a specific example they have considered the Coulomb potential $V(r) = -\beta/r$, for which the relativistic wave equations are exactly soluble in N space dimensions.[10] For a spin-zero particle, for an example, the exact energy (in units $\hbar = c = 1$) is

$$E = m[1+4\beta^2\{2n-2l+1 + \sqrt{(k-2)^2-4\beta^2}\}^{-2}]^{-1/2}, \tag{1}$$

which, when expanded in powers of $1/k$, becomes, for the ground state,

$$E = m\left[1-\sum_{n=1} C_n\frac{2\beta^{2n}}{k^{2n}}\right], \tag{2}$$

where the first three coefficients C_1, C_2, and C_3 are given by

$$\begin{aligned} C_1 &= 1+\frac{2}{k}+\frac{3}{k^2}+\frac{4}{k^3}+\cdots,\\ C_2 &= 1+\frac{8}{k}+\frac{34}{k^2}+\frac{108}{k^3}+\cdots,\\ C_3 &= 2\left(1+\frac{12}{k}+\frac{81}{k^2}+\frac{400}{k^3}+\cdots\right). \end{aligned} \tag{3}$$

Obviously the relativistic corrections are nonleading in $1/k$. For most potentials, however, the relativistic wave equations are not exactly soluble and one, therefore, has to resort to some approximation scheme. In the present paper we make an attempt in this direction. We propose a large-N iterative procedure for the solution of the Klein–Gordon equation to obtain the energy eigenvalue of a spin-0 particle in a potential $V(r)$. For the Coulomb potential the method is found to generate the exact series.

II. LARGE-N ITERATIVE PROCEDURE

The N-dimensional Klein–Gordon equation (in units $\hbar = c = 1$) for a scalar particle of mass m moving in a Lorentz vector potential, whose only surviving component is the fourth component $V(r)$, is given by

$$[\nabla^2 + (E-V(r))^2 - m^2]\Psi(\mathbf{r}) = 0, \tag{4}$$

where $\mathbf{r}$ is an N-dimensional vector of magnitude r and ∇^2 can be written as

$$\nabla^2 = \frac{\partial^2}{\partial r^2}+\frac{N-1}{r}\frac{\partial}{\partial r}-\frac{L^2}{r^2}, \tag{5}$$

L being the angular momentum operator in N dimensions. Substituting in (4)

$$\Psi(\mathbf{r}) = R(r)Y_{lm}(\Omega), \tag{6}$$

where $R(r)$ is a function of r and $Y_{lm}(\Omega)$ is an eigenfunction of L^2 belonging to the eigenvalue $l(l+N-2)$, we get

$$\left\{\frac{\partial^2}{\partial r^2}+\frac{N-1}{r}\frac{\partial}{\partial r}-\frac{l(l+N-2)}{r^2} + (E-V(r))^2-m^2\right\}R(r) = 0, \tag{7}$$

which, on substituting $R(r) = r^{-(N-1)/2}u(r)$, reduces to

$$\left[-\frac{d^2}{dr^2}+\frac{(k-1)(k-3)}{4r^2} - \{(E-V(r))^2-m^2\}\right]u(r) = 0, \tag{8}$$

where $k = N+2l$. In the large-k limit ($N\to\infty$), the energy to leading order is given by

$$E_\infty = V(r_0) + m[1 + k^2/4m^2r_0^2]^{1/2}, \tag{9}$$

where r_0 is the value of r at which the wave function (in the limit $k \to \infty$) has a δ-function peak and is to be determined from

$$r_0^3 V'(r_0)(1 + k^2/4m^2r_0^2)^{1/2} = k^2/4m, \tag{10}$$

where $V'(r_0) = [dV(r)/dr]|_{r=r_0}$.

To evaluate the finite-N corrections we rewrite Eq. (8) as

$$\Big[-\frac{d^2}{dr^2} + \frac{(k-1)(k-3)}{4r^2} - \{(E_a - V(r))^2 - m^2\} + 2(E - E_a)(V(r) - V(r_0))\Big]u(r) = [(E - V(r_0))^2 - (E_a - V(r_0))^2]u(r), \tag{11}$$

where E_a is some approximate solution for E. Now the simplest approximation consists in putting $E_a = E_\infty$ and neglecting the term $2(E - E_\infty)(V(r) - V(r_0))$. Then denoting the energy in this approximation by $E^{(1)}$ and the wave function by $u^{(1)}(r)$, we get from (11)

$$\Big[-\frac{d^2}{dr^2} + \frac{(k-1)(k-3)}{4r^2} - \{(E_\infty - V(r))^2 - m^2\}\Big]u^{(1)}(r) = \mathscr{E}_1 u^{(1)}(r), \tag{12}$$

where

$$\mathscr{E}_1 = (E^{(1)} - V(r_0))^2 - (E_\infty - V(r_0))^2. \tag{13}$$

Now for the ground state wave function (which is nodeless), we assume, following Mlodinow and Shatz,[2]

$$u^{(1)}(r(x)) = \exp[\Phi_1(x)], \tag{14}$$

where $x = r - r_0$. Equation (12) then reduces to the Riccati equation

$$-(\Phi_1''(x) + \Phi_1'^2(x)) + k^2 V_{\rm eff}^{(1)}(x) + \frac{k}{r^2(x)} + \frac{3}{4r^2(x)} = \mathscr{E}_1, \tag{15}$$

where

$$\Phi_1'(x) = \frac{d\Phi_1(x)}{dx}, \quad \Phi_1''(x) = \frac{d^2\Phi_1(x)}{dx^2},$$

and

$$V_{\rm eff}^{(1)}(x) = \left(\frac{1}{4r^2(x)} + \frac{m}{k^2}\right) - \frac{(E_\infty - V(r(x))^2}{k^2}. \tag{16}$$

Next substituting the expansions

$$\mathscr{E}_1 = \sum_{n=-1}^{\infty} \mathscr{E}_1^{(n)} k^{-n} \tag{17}$$

and

$$\Phi_1'(x) = \sum_{n=-1}^{\infty} \Phi_1^{(n)}(x) k^{-n} \tag{18}$$

in (15) and equating the terms of same order in k we generate the following recurrence relations:

$$\begin{aligned}
&\Phi_1^{(-1)}(x) = -(V_{\rm eff}^{(1)}(x))^{1/2},\\
&2(V_{\rm eff}^{1}(x))^{1/2}\Phi_1^{(0)}(x) = \mathscr{E}_1^{(-1)} + \Phi_1^{(-1)'}(x) + r^{-2}(x),\\
&2(V_{\rm eff}^{(1)}(x))^{1/2}\Phi_1^{(1)}(x) = \mathscr{E}_1^{(0)} + \Phi_1^{(0)'}(x) - \tfrac{3}{4}r^{-2}(x),\\
&2(V_{\rm eff}^{(1)}(x))^{1/2}\Phi_1^{(n+1)}(x) = \mathscr{E}_1^{(n)} + \Phi_1^{(n)'}(x) + \sum_{m=0}^{n} \Phi_1^{m}(x)\Phi_1^{(n-m)}(x),\\
&n > 0.
\end{aligned} \tag{19}$$

Since $V_{\rm eff}^{(1)}(0) = 0$, we have from (19)

$$\begin{aligned}
\mathscr{E}_1^{(-1)} &= -\Phi_1^{(-1)'}(0) - r_0^{-2},\\
\mathscr{E}_1^{(0)} &= -\Phi_1^{(0)'}(0) + \tfrac{3}{4}r_0^{-2},\\
\mathscr{E}_1^{(n)} &= -\Phi_1^{(n)'}(0) - \sum_{m=0}^{n} \Phi_1^{(m)}(0)\Phi_1^{(n-m)}(0), \quad n > 0.
\end{aligned} \tag{20}$$

Then $E^{(1)}$ is obtained from

$$E^{(1)} = V(r_0) + [(E_\infty - V(r_0))^2 + \mathscr{E}_1]^{1/2}. \tag{21}$$

In the next improved approximation we set $E_a = E^{(1)}$ in (11) and drop the term $2(E - E^{(1)})(V(r) - V(r_0))$. Let us denote the energy in this approximation by $E^{(2)}$ and the wave function by $u^{(2)}(r)$. Thus we obtain

$$\Big[-\frac{d^2}{dr^2} + \frac{(k-1)(k-3)}{4r^2} - \{(E^{(1)} - V(r))^2 - m^2\}\Big]u^{(2)}(r) = [(E^{(2)} - V(r_0))^2 - (E^{(1)} - V(r_0))^2]u^{(2)}(r), \tag{22}$$

which can be solved again by the above method of $1/N$ expansion by transforming it to the Riccati equation. After $E^{(2)}$ is determined, E_a should be replaced by $E^{(2)}$ in (11) and then, ignoring the term $2(E - E^{(2)})(V(r) - V(r_0))$, the resulting equation may be solved to obtain a more improved approximate solution $E^{(3)}$ for the energy E. This iterative procedure obviously can be continued to obtain any desired order of accuracy. In the next section we apply this method to the Coulomb potential for which the method yields exact results.

III. APPLICATION TO THE COULOMB POTENTIAL

In the case of the Coulomb potential, $V(r) = -\beta/r$, we have

$$r_0 = (k^2/4m\beta)(1 - 4\beta^2/k^2)^{1/2} \tag{23}$$

and

$$E_\infty = m(1 - 4\beta^2/k^2)^{1/2}. \tag{24}$$

Then $V_{\rm eff}^{(1)}(x)$ is given by

$$V_{\rm eff}^{(1)}(x) = (4m^2\beta^2/k^4)(1 - r_0/r)^2, \tag{25}$$

and, using (16), (18), and (19), we obain

$$\mathscr{E}_1 = m^2\Big[-\frac{4\beta^2}{k^2}\Big(\frac{2}{k} + \frac{3}{k^2} + \frac{4}{k^3} + \cdots\Big) - \frac{\beta^4}{k^4}\Big(\frac{48}{k} + \frac{208}{k^2} + \frac{608}{k^3} + \cdots\Big) - \cdots\Big], \tag{26}$$

so that $E^{(1)}$ becomes

$$E^{(1)} = m\left[1 - \sum_{n=1}^{\infty} D_n \frac{2\beta^{2n}}{k^{2n}}\right], \tag{27}$$

with

$$D_1 = 1 + \frac{2}{k} + \frac{3}{k^2} + \frac{4}{k^3} + \cdots,$$
$$D_2 = 1 + \frac{8}{k} + \frac{50}{k^2} + \frac{156}{k^3} + \cdots. \tag{28}$$

and so on. Thus the nonrelativistic part of the Coulomb ground state energy is exactly recovered even in the simplest approximation, the relativistic corrections remaining inexact, however. To obtain, therefore, a more accurate solution, we now consider (22), which can be written as

$$\left[-\frac{d^2}{dr^2} + \frac{(k-1)(k-3)}{4r^2} - \{(E_\infty - V(r))^2 - m^2\} + 2(E^{(1)} - E_\infty)(V(r) - V(r_0))\right]u^{(2)}(r) = \mathscr{E}_2 u^{(2)}(r), \tag{29}$$

where

$$\mathscr{E}_2 = (E^{(2)} - V(r_0))^2 - (E_\infty - V(r_0))^2. \tag{30}$$

Substituting as before,

$$u^{(2)}(r) = \exp[\Phi_2(x)], \tag{31}$$

we get from (29)

$$(\Phi_2''(x) + \Phi_2'^2(x)) + k^2 V_{\text{eff}}^{(2)}(x) - kr^{-2}(x) - \tfrac{3}{4}r^{-2}(x) = \tilde{\mathscr{E}}_2, \tag{32}$$

where

$$V_{\text{eff}}^{(2)}(x) = \frac{1}{4r^2} + \frac{m}{k^2} - \frac{(E_\infty - V(r))^2}{k^2} + \frac{2(E^{(1)} - E_\infty)(V(r) - V(r_0))}{k^2} + \frac{(E^{(1)} - E_\infty)^2}{4m^2 r_0^2}, \tag{33}$$

and

$$\tilde{\mathscr{E}}_2 = \mathscr{E}_2 + \frac{(E^{(1)} - E_\infty)^2 k^2}{4m^2 r_0^2}. \tag{34}$$

Employing the expansions

$$\tilde{\mathscr{E}}_2 = \sum_{n=-1}^{\infty} \tilde{\mathscr{E}}_2^{(n)} k^{-n} \tag{35}$$

and

$$\Phi_2'(x) = \sum_{n=-1}^{\infty} \Phi_2^{(n)}(x) k^{-n}, \tag{36}$$

we end up with a set of recurrence relations similar to (18), which are now to be solved utilizing the condition that

$$V_{\text{eff}}^{(2)}(x) = \frac{4m^2\beta^2}{k^2}\left(1 - \frac{r_0}{r} + \frac{(E^{(1)} - E_\infty)k^2}{4m^2\beta r_0}\right)^2 \tag{33'}$$

vanishes at

$$r = \bar{r} = \bar{x} + r_0 = \frac{4m^2\beta r_0^2}{4m^2\beta r_0 + (E^{(1)} - E_\infty)k^2}. \tag{37}$$

The energy $E^{(2)}$ is finally given by

$$E^{(2)} = V(r_0) + \left[(E_\infty - V(r_0))^2 - \frac{(E^{(1)} - E_\infty)^2 k^2}{4m^2 r_0^2} + \tilde{\mathscr{E}}_2\right]^{1/2} \tag{38}$$

$$= m\left[1 - \frac{2\beta^2}{k^2}\left(1 + \frac{2}{k} + \frac{3}{k^2} + \frac{4}{k^3} + \cdots\right) - \frac{2\beta^4}{k^4}\left(1 + \frac{8}{k} + \frac{34}{k^2} + \frac{108}{k^3} + \cdots\right) - \cdots\right]. \tag{39}$$

Comparison of (39) with (2) reveals that $E^{(2)}$ is exact to the order considered. For higher-order relativistic corrections, however, one has to make further iterations.

In the case of higher excited states for which the wave functions $u(r)$ are to be chosen to have the right number of nodes, the above method of $1/N$ expansion becomes extremely cumbersome. Alternatively one can employ the perturbed oscillator method[11] to which we now turn.

IV. THE PERTURBED OSCILLATOR METHOD

In the perturbed oscillator method we write Eq. (12) as

$$\left[-\frac{1}{2m}\frac{d^2}{dr^2} + \frac{k^2}{8mr^2} - \frac{k}{2mr^2} + \frac{3}{8mr^2} + k^2 f_1(r)\right]u^{(1)}(r) = (\mathscr{E}_1/2m)u(r), \tag{40}$$

where

$$f_1(r) = [m^2 - (E_\infty - V(r))^2]/2mk^2. \tag{41}$$

Next we define $y = (\sqrt{k}/r_0)(r - r_0)$ and expand the r-dependent terms in (40) around $y = 0$ to obtain an effective one-dimensional perturbed (nonrelativistic) harmonic oscillator equation[11]

$$\left[-\frac{1}{2m}\frac{d^2}{dy^2} + \frac{1}{2}m^2\omega^{(1)^2}y^2 + \varepsilon_0^{(1)} + \hat{V}^{(1)}(y)\right]\varphi^{(1)}(y) = \lambda^{(1)}\varphi^{(1)}(y), \tag{42}$$

where

$$\omega^{(1)} = \left[\frac{3}{4m^2} + \frac{r_0^4 f_1''(r_0)}{m}\right]^{1/2}, \tag{43}$$

$$\varepsilon_0^{(1)} = \frac{k}{8m} - \frac{1}{2m} + \frac{3}{8mk} + r_0^2 k f(r_0), \tag{44}$$

$$\lambda^{(1)} = (r_0^2/2mk)\mathscr{E}_1, \tag{45}$$

and

$$\hat{V}^{(1)}(y) = \frac{1}{k^{1/2}}(\varepsilon_1^{(1)}y + \varepsilon_3^{(1)}y^3) + \frac{1}{k}(\varepsilon_2^{(1)}y^2 + \varepsilon_4^{(1)}y^4) + \frac{1}{k^{3/2}}(\delta_1^{(1)}y + \delta_3^{(1)}y^3 + \delta_5^{(1)}y^5) + \frac{1}{k^2}(\delta_2^{(1)}y^2 + \delta_4^{(1)}y^4 + \delta_6^{(1)}y^6) + \cdots, \tag{46}$$

$$\varepsilon_1^{(1)}=\frac{1}{m},\quad \varepsilon_2^{(1)}=-\frac{3}{2m},\quad \varepsilon_3^{(1)}=\left(\frac{r_0^5 f_1'''(r_0)}{6}-\frac{1}{2m}\right),\quad \varepsilon_4^{(1)}=\left(\frac{r_0^6 f_1''''(r_0)}{24}+\frac{5}{2m}\right);$$

$$\delta_1^{(1)}=-\frac{3}{4m},\quad \delta_2^{(1)}=\frac{9}{8m},\quad \delta_3^{(1)}=\frac{2}{m},\quad \delta_4^{(1)}=-\frac{5}{2m},\quad \delta_5^{(1)}=\left(\frac{r_0^7 f_1'''''(r_0)}{120}-\frac{3}{4m}\right),\quad \delta_6^{(1)}=\left(\frac{r_0^8 f_1''''''(r_0)}{720}+\frac{7}{8m}\right). \tag{47}$$

Now applying the fourth-order perturbation theory to (42) and arranging terms of the energy in powers of $1/k$, we obtain

$$\begin{aligned}
E^{(1)} = V(r_0) + m\Bigg[&\left(1+\frac{k^2}{4m^2r_0^2}\right)+\frac{2}{m}\Bigg\{k^2\left(\frac{1}{8mr_0^2}+f_1(r_0)\right)+k\left(\frac{(n+\frac{1}{2})\omega^{(1)}}{r_0^2}-\frac{1}{2mr_0^2}\right)\\
&+\Bigg(\frac{3}{8mr_0^2}+\frac{(1+2n)}{r_0^2}\bar\varepsilon_2^{(1)}+\frac{3(1+2n+2n^2)}{r_0^2}\bar\varepsilon_4^{(1)}\\
&-\frac{1}{\omega^{(1)}r_0^2}\left[\bar\varepsilon_1^{(1)^2}+6(1+2n)\bar\varepsilon_1^{(1)}\bar\varepsilon_3^{(1)}+(11+30n+30n^2)\bar\varepsilon_3^{(1)^2}\right]\Bigg)\\
&+\frac{1}{k}\Bigg(\frac{1}{r_0^2}\left[(1+2n)\bar\delta_2^{(1)}+3(1+2n+2n^2)\bar\delta_4^{(1)}+5(3+8n+6n^2+4n^3)\bar\delta_6^{(1)}\right]\\
&-\frac{1}{\omega^{(1)}r_0^2}\Big[(1+2n)\bar\varepsilon_2^{(1)^2}+12(1+2n+2n^2)\bar\varepsilon_2^{(1)}\bar\varepsilon_4^{(1)}\\
&+2(21+59n+51n^2+34n^3)\bar\varepsilon_4^{(1)^2}+2\bar\varepsilon_1^{(1)}\bar\delta_1^{(1)}+6(1+2n)\bar\varepsilon_1^{(1)}\bar\delta_3^{(1)}\\
&+30(1+2n+2n^2)\bar\varepsilon_1^{(1)}\bar\delta_5^{(1)}+6(1+2n)\bar\varepsilon_3^{(1)}\bar\delta_1^{(1)}\\
&+2(11+30n+30n^2)\bar\varepsilon_3^{(1)}\bar\delta_3^{(1)}+10(13+40n+42n^2+28n^3)\bar\varepsilon_3^{(1)}\bar\delta_5^{(1)}\Big]\\
&+\frac{1}{\omega^{(1)^2}r_0^2}\Big[4\bar\varepsilon_1^{(1)^2}\bar\varepsilon_2^{(1)}+36(1+2n)\bar\varepsilon_1^{(1)}\bar\varepsilon_2^{(1)}\bar\varepsilon_3^{(1)}+8(11+30n+30n^2)\bar\varepsilon_2^{(1)}\bar\varepsilon_3^{(1)^2}\\
&+24(1+2n)\bar\varepsilon_1^{(1)^2}\bar\varepsilon_4^{(1)}+8(31+78n+78n^2)\bar\varepsilon_1^{(1)}\bar\varepsilon_3^{(1)}\bar\varepsilon_4^{(1)}+12(57+189n+225n^2+150n^3)\bar\varepsilon_3^{(1)^2}\bar\varepsilon_4^{(1)}\Big]\\
&-\frac{1}{\omega^{(1)^3}r_0^2}\Big[8\bar\varepsilon_1^{(1)^3}\bar\varepsilon_3^{(1)}+108(1+2n)\bar\varepsilon_1^{(1)^2}\bar\varepsilon_3^{(1)^2}+48(11+30n+30n^2)\bar\varepsilon_1^{(1)}\bar\varepsilon_3^{(1)^3}\\
&+30(31+109n+141n^2+94n^3)\bar\varepsilon_3^{(1)^4}\Big]\Bigg)+\cdots\Bigg\}\Bigg]^{1/2},
\end{aligned} \tag{48}$$

where

$$\bar\varepsilon_j^{(1)}=\varepsilon_j^{(1)}/[2m\omega^{(1)}]^{j/2},\quad \bar\delta_j^{(1)}=\delta_j/[2m\omega^{(1)}]^{j/2}. \tag{49}$$

After $E^{(1)}$ is obtained, we proceed to solve Eq. (29) for $E^{(2)}$. Defining

$$f_2(r)=\frac{m^2-(E_\infty-V(r))^2}{2mk^2}+\frac{(E^{(1)}-E_\infty)}{mk^2}(V(r)-V(r_0)), \tag{50}$$

we get from (29)

$$\left[-\frac{1}{2m}\frac{d^2}{dr^2}+\frac{k^2}{8mr^2}-\frac{k}{2mr^2}+\frac{3}{8mr^2}+k^2f_2(r)\right]u^{(2)}(r)=\frac{\mathscr{E}_2}{2m}u^{(2)}(r), \tag{51}$$

which looks similar to (40) and hence it can be solved by the fourth-order perturbation theory by reducing it to the perturbed oscillator equation

$$\left[-\frac{1}{2m}\frac{d^2}{dy^2}+\frac{1}{2}m^2\omega^{(2)^2}y^2+\varepsilon_0^{(2)}+\hat V^{(2)}(y)\right]\varphi^{(2)}(y)=\lambda^{(2)}\varphi^{(2)}(y), \tag{52}$$

where

$$\omega^{(2)}=[3/4m^2+r_0^4f_2''(r_0)/m]^{1/2}, \tag{53}$$

$$\varepsilon_0^{(2)} = \frac{k}{8m} - \frac{1}{2m} + \frac{3}{8mk} + r_0^2 k f_2(r_0)\,, \tag{54}$$

$$\lambda^{(2)} = (r_0^2/2mk)\mathscr{E}_2\,, \tag{55}$$

and

$$\hat{V}^{(2)}(y) = \frac{1}{k^{1/2}}(\varepsilon_1^{(2)}y + \varepsilon_3^{(2)}y^3) + \frac{1}{k}(\varepsilon_2^{(2)}y^2 + \varepsilon_4^{(2)}y^4)$$

$$+ \frac{1}{k^{3/2}}(\delta_1^{(2)}y + \delta_3^{(2)}y^3 + \delta_5^{(2)}y^5) + \frac{1}{k^2}(\delta_2^{(2)}y^2 + \delta_4^{(2)}y^4 + \delta_6^{(2)}y^6) + \cdots\,, \tag{56}$$

with

$$\varepsilon_1^{(2)} = \left[\frac{1}{m} + \frac{r_0^3 V'(r_0)}{mk}(E^{(1)} - E_\infty)\right],\quad \varepsilon_2^{(2)} = -\frac{3}{2m},\quad \varepsilon_3^{(2)} = \left(\frac{r_0^5 f_2'''(r_0)}{6} - \frac{1}{2m}\right),\quad \varepsilon_4^{(2)} = \left(\frac{r_0^6 f_2''''(r_0)}{24} + \frac{5}{8m}\right);$$

$$\delta_1^{(2)} = -\frac{3}{4m},\quad \delta_2^{(2)} = \frac{9}{8m},\quad \delta_3^{(2)} = \frac{2}{m},\quad \delta_4^{(2)} = -\frac{5}{2m},\quad \delta_5^{(2)} = \left(\frac{r_0^7 f_2'''''(r_0)}{120} - \frac{3}{4m}\right),\quad \delta_6^{(2)} = \left(\frac{r_0^8 f_2''''''(r_0)}{720} + \frac{7}{8m}\right). \tag{57}$$

After $E^{(2)}$ is calculated one has to solve

$$\left[-\frac{1}{2m}\frac{d^2}{dr^2} + \frac{k^2}{8mr^2} - \frac{k}{2mr^2} + \frac{3}{8mr^2} + k^2 f_3(r)\right]u^{(3)}(r) = \frac{\mathscr{E}_3}{2m}u^{(3)}(r)\,, \tag{58}$$

where

$$f_3(r) = \frac{m^2 - (E_\infty - V(r))^2}{2mk^2} + \frac{(E^{(2)} - E_\infty)}{mk^2}(V(r) - V(r_0)) \tag{59}$$

and

$$\mathscr{E}_3 = (E^{(3)} - V(r_0))^2 - (E_\infty - V(r_0))^2 \tag{60}$$

to obtain a more improved solution $E^{(3)}$ and similarly the process can be continued to any order.

The method delineated above can be readily applied to any spherically symmetric potential to obtain its energy spectrum. For the Coulomb potential $V(r) = -\beta/r$, the results of the first few iterations are

$$E^{(1)} = m\left[1 - \frac{2\beta^2}{k^2}\left(1 + \frac{2}{k} + \frac{3}{k^2} + \cdots\right) - \frac{2\beta^4}{k^4}\left(1 + \frac{8}{k} + \frac{50}{k^2} + \cdots\right) - \cdots\right], \tag{61}$$

$$E^{(2)} = m\left[1 - \frac{2\beta^2}{k^2}\left(1 + \frac{2}{k} + \frac{3}{k^2} + \cdots\right) - \frac{2\beta^4}{k^4}\left(1 + \frac{8}{k} + \frac{34}{k^2} + \cdots\right) - \cdots\right], \tag{62}$$

$$E^{(3)} = m\left[1 - \frac{2\beta^2}{k^2}\left(1 + \frac{2}{k} + \frac{3}{k^2} + \cdots\right) - \frac{2\beta^4}{k^4}\left(1 + \frac{8}{k} + \frac{34}{k^2} + \cdots\right) - \cdots\right]. \tag{63}$$

Thus in the case of the Coulomb potential one may stop at $E^{(2)}$ for the lowest-order relativistic correction.

V. CONCLUSION

We have developed in this paper an iterative large-N scheme by extending the method of Mlodinow and Shatz[2] and the perturbed oscillator method[11] to obtain the energy eigenvalue of a scalar particle moving in a spherically symmetric potential. For the Coulomb potential in particular, the method looks quite attractive, for it yields the exact energy series, at least to the order considered in this paper. Finally we remark that our procedure may be useful in the qualitative studies of quark-confining potentials.

[1] L. Mlodinow and N. Papanicolaou, Ann. Phys. (NY) **128**, 314 (1980); **131**, 1 (1981).

[2] L. Mlodinow and M. Shatz, J. Math. Phys. **25**, 943 (1984).

[3] A. Chatterjee, J. Phys. A: Math. Gen. **18**, 1193 (1985); "1/N expansion for the Yukawa potential revisited. II," J. Phys. A: Math. Gen. (in press).

[4] E. Witten, Nucl. Phys. B **160**, 57 (1979).

[5] S. Coleman, SLAC preprint Pub-2484, March 1980, Erice Lectures, 1979.

[6] S. Ma, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1976), Vol. 6, p. 250.

[7] E. Brezin, J. C. Le Guillou, and J. Zinn-Justin, in Ref. 6, p. 127.

[8] A. Chatterjee, J. Phys. A: Math. Gen. **18**, 735 (1985).

[9] J. L. Miramontes and C. Pajares, Nuovo Cimento B **84**, 10 (1984).

[10] M. M. Neito, Am. J. Phys. **47**, 1067 (1979).

[11] U. Sukhatme and T. Imbo, Phys. Rev. D **28**, 418 (1983).

Conformal transformations and viscous fluids in general relativity

J. Carot and Ll. Mas
Dpt. Fisica Teorica, Universitat Illes Balears, Ctra. Valldemossa, km 7.5, 07071 Palma de Mallorca, Spain

(Received 6 January 1986; accepted for publication 30 April 1986)

It is shown that viscous fluid solutions can be obtained by performing conformal transformations of vacuum solutions of Einstein's field equations. The solutions obtained by such a procedure can be matched, under certain conditions, to their respective original vacuum metrics.

I. INTRODUCTION

Anisotropic fluids have proved to be of great interest as alternative (and more realistic models) to perfect fluids when describing some physical situations for consideration in general relativity such as nonstatic interiors of stars,[1–3] evolution of radiating spheres,[4,5] etc. Viscous fluids[6–9] are a special type of anisotropic fluids that possess the relevant characteristic that the anisotropic pressure tensor is proportional to the shear tensor of the velocity field of the fluid. Thus, the constant of proportionality is a multiple of the kinematical viscosity. These fluids arouse special interest because they explain relativistic dissipative processes in situations near the thermodynamic equilibrium[6] and have been used to describe neutron stars in certain density ranges[10] and likewise in cosmology.[11] From a more theoretical point of view, the possibility of viscous fluids generating some already known exact solutions has also been studied in FRW models,[12] electromagnetic field solutions,[13,14] scalar fields,[15] etc.

In the present paper a method is described for generating viscous fluid solutions starting from a vacuum solution and then carrying out a conformal transformation.[16] In Sec. II a brief account of the theory of conformal transformations is given (for further details see the references cited therein). Moreover, it is shown how a conformal transformation of a vacuum solution generates (through Einstein's field equations) a matter solution by giving the stress-energy tensor corresponding to the new solution in terms of the function generating the conformal transformation. Section III is dedicated to the study of anisotropic and viscous fluids specifying the general form of their stress-energy tensors and the definition and properties of the magnitudes taking part in their composition. In Sec. IV the results of the two preceding sections are applied to identify the stress-energy tensor corresponding to the new solution given in Sec. II. The possibility of this tensor corresponding to a viscous fluid with heat conduction is given, with negative expansion taking place in most of the cases; this is a collapsing fluid. We have specified under which conditions such a viscous fluid reduces to a perfect one and briefly discussed the matching problem between both metrics: the new solution and the vacuum solution from which it was derived. Finally, in Sec. V, the former results are applied to a particular case, namely, the exterior Schwarzschild solution, and a "naive" model of a spherically symmetric collapsing star is obtained, showing viscosity, anisotropic pressures, and heat conduction. This model satisfies energy conditions[17] and the pressure and the density show a reasonable behavior. However, it is not our intention to put forward this model as a realistic one, it is only used to illustrate the general results formerly obtained. Further models are obtainable likewise by performing some slight changes in the present one. Moreover, this method can be applied to other vacuum solutions, such as Minkowski flat space, as is indicated in Sec. V.

II. CONFORMAL TRANSFORMATIONS IN GENERAL RELATIVITY

As is well known,[16,17] conformal transformations constitute a special type of map of metric spaces consisting of a dilatation (or contraction) of all lengths by a common factor depending on the point of the space:

$$\tilde{g}_{ab} = e^{2U} g_{ab}, \quad U = U(x^c). \tag{1}$$

The most important property satisfied by conformal transformations (1) is that they keep the components of the Weyl's conformal tensor unchanged, and therefore the Bel–Petrov type, too. Another important property that will be useful later is that the isometry group of the original metric g_{ab} becomes a conformal group of motions of the new metric $\tilde{g}_{ab}$. Some more interesting properties of conformal transformations may be found in Ref. 16.

The Einstein tensors corresponding to the metrics g and $\tilde{g}$ in the case of four-dimensional spaces are connected by

$$\begin{aligned}\widetilde{G}_{ab} = G_{ab} &+ 2(U_a U_b - \tfrac{1}{2} U^c U_c g_{ab}) \\ &+ 2(U^c{}_{;c} + U^c U_c) g_{ab} - 2U_{a;b},\end{aligned} \tag{2}$$

where the covariant derivatives and contractions are calculated with the metric g_{ab}, and U_a stands for $U_{,a}$. The former expression written in terms of the new metric $\tilde{g}_{ab}$ (namely, contractions and covariant derivatives performed with $\tilde{g}_{ab}$) reads

$$\widetilde{G}_{ab} = G_{ab} - 2U_a U_b - 2U_{a;b} + (2U^c{}_{;c} - U^c U_c)\tilde{g}_{ab}. \tag{3}$$

If g_{ab} is chosen to be a vacuum solution its associated Einstein tensor is zero, and then, by using Einstein's field equations $G_{ab} = T_{ab}$, the second member of (3) with $G_{ab} = 0$ may be interpreted as the stress-energy tensor of some distribution of matter,

$$\widetilde{T}_{ab} = -2U_a U_b - 2U_{a;b} - (U^c U_c - 2U^c{}_{;c})\tilde{g}_{ab}. \tag{4}$$

The identification of the material distribution described by (4) will be discussed in Sec. IV.

III. ANISOTROPIC AND VISCOUS FLUIDS

The stress-energy tensor describing an anisotropic fluid with heat conduction[9] can be written as

$$T_{ab} = \rho u_a u_b + p h_{ab} + \pi_{ab} + u_a q_b + q_a u_b, \quad (5)$$

where ρ is the energy density, p is the thermodynamic pressure, u_a is the velocity field of the fluid ($u^a u_a = -1$), q^a is the heat conduction vector ($q^a q_a > 0$ and $q^a u_a = 0$), π_{ab} is the anisotropic pressure tensor ($g^{ab}\pi_{ab} = \pi_{ab} u^b = 0$), and h_{ab} is the orthogonal projector to the velocity field $h_{ab} = g_{ab} + u_a u_b$. Note that π_{ab} and $q_a u_b + u_a q_b$ are dissipative terms that do not appear in the stress-energy tensor of a perfect fluid.

A particular case of an anisotropic fluid is that of a viscous fluid, characterized by an anisotropic pressure tensor π_{ab} proportional to the shear tensor σ_{ab} of the velocity field u_a,[6–9,18]

$$\pi_{ab} = -2\eta\sigma_{ab}, \quad (6)$$

$$\sigma_{ab} = \tfrac{1}{2}\{u_{a;b} + u_{b;a} + \dot{u}_a u_b + u_a \dot{u}_b\} - (\theta/3)h_{ab}, \quad (7)$$

where η is the kinematical viscosity coefficient ($\eta > 0$), $\dot{u}_a$ is the acceleration of the fluid ($\dot{u}_a = u_{a;b}u^b$), and θ is the expansion of the fluid ($\theta = u^c{}_{;c}$). (For the physical sense of θ, σ_{ab}, etc. see, for instance, Ref. 9.) Therefore, the stress-energy tensor of a viscous fluid with heat conduction is

$$T_{ab} = \rho u_a u_b + p h_{ab} - 2\eta\sigma_{ab} + u_a q_b + q_a u_b, \quad (8)$$

which constitutes the expression of the stress-energy tensor for a relativistic dissipative fluid in the framework of the standard theory for such fluid mechanics developed by Eckart[6,19] and by Landau and Lifshitz,[7] and it establishes the relativistic equivalent to the Navier–Stokes theory of Newtonian fluid mechanics.

Note that the quantities characterizing an anisotropic fluid,[5] namely, ρ, p, q_a, and π_{ab}, can be written in an "intrinsic" manner if the velocity field u^a is known (this is completely general for any symmetric second-rank tensor; that is, given a timelike vector one can always perform such a decomposition of the given tensor):

$$\rho = T_{ab}u^a u^b, \quad p = \tfrac{1}{3}h^{ab}T_{ab},$$
$$q_a = -h_a{}^b T_{bc}u^c, \quad \pi_{ab} = h_a{}^c h_b{}^d (T_{cd} - p g_{cd}). \quad (9)$$

IV. IDENTIFICATION OF T_{ab}

The purpose of this section is to show that the stress-energy tensor given by (4) may be interpreted as the tensor of a viscous fluid with heat conduction, provided that the function U in (1) satisfies certain requirements. To do this we shall substitute the tensor (4) in the set of expressions (9), and we shall take as the velocity field of the supposed fluid the normalized gradient of the function U; then, the only *a priori* restriction on U will be that its gradient is timelike:

$$u_a \doteq (1/\lambda)U_a, \quad \lambda \doteq (-U_a U^a)^{1/2}. \quad (10)$$

Equation (4) now reads

$$T_{ab} = -2\lambda^2 u_a u_b - 2(\lambda_{(b}u_{a)} + \lambda u_{(a;b)}) + (\lambda^2 + 2\lambda\theta + 2\dot{\lambda})\tilde{g}_{ab}, \quad (11)$$

where λ_b stands for $\lambda_{,b}$ and $\dot{\lambda}$ stands for the derivative of λ in the direction of the velocity field $\dot{\lambda} = \lambda_a u^a$.

The density, pressure, heat conduction, and anisotropic pressure tensor calculated through (9) are

$$\rho = -(3\lambda^2 + 2\lambda\theta), \quad (12)$$

$$p = \lambda^2 + 2\dot{\lambda} + \tfrac{4}{3}\lambda\theta, \quad (13)$$

$$q_a = 2\lambda\dot{u}_a, \quad (14)$$

$$\pi_{ab} = -2\lambda\{u_{(a;b)} + \dot{u}_{(a}u_{b)} - (\theta/3)h_{ab}\} = -2\lambda\sigma_{ab}. \quad (15)$$

From (15) we see that the anisotropic pressure tensor is proportional to the shear of the velocity field, and this also permits us to identify λ with the viscosity coefficient, that is,

$$\eta = \lambda. \quad (16)$$

Therefore, the fluid under consideration is not only anisotropic, but, more specifically, viscous.

Some other properties can be worked out from expressions (12)–(15); for instance, the requirement $\rho > 0$ (dominant energy condition[23]) implies $\theta < 0$, and therefore

$$|\theta| > \tfrac{3}{2}\lambda. \quad (17)$$

The fluid so obtained is irrotational ($\omega_{ab} = 0$) since its velocity field is hypersurface orthogonal; it is collapsing (negative expansion); and, provided that on the surface $U = 0$ the derivatives of U and the pressure p vanish, it can be matched continuously to the vacuum solution from which it derives.

If the dissipative terms in the stress-energy tensor (8) vanish, i.e., $-2\eta\sigma_{ab} + q_a u_b + u_a q_b = 0$, the viscous fluid then becomes a perfect fluid. The necessary and sufficient condition for this to take place is that both q_a and σ_{ab} vanish separately. In the present case, it is easy to see what these conditions imply:

$$\text{(i) } q_a = 0 \Leftrightarrow \dot{u}_a = 0, \quad \text{geodesic fluid,}$$
$$\text{(ii) } \sigma_{ab} = 0 \Leftrightarrow u_{a;b} = 1/3\theta h_{ab}, \quad \text{shear-free.} \quad (18)$$

There is still another restriction upon the velocity field u_a, and consequently upon the gradient of the function U: the component u^0 must be positive in order to preserve the orientation towards the future of the temporal coordinate. This implies $U_t < 0$ in the former expressions; otherwise, if $U_t > 0$, we should define the velocity field as

$$u_a \doteq -(1/\lambda)U_a \quad (\lambda > 0), \quad (19)$$

and then the pressure, density, heat conduction, and anisotropic pressure tensor become

$$\rho = -3\lambda^2 + 2\lambda\theta, \quad (20)$$

$$p = \lambda^2 - 2\dot{\lambda} - \tfrac{4}{3}\lambda\theta, \quad (21)$$

$$q_a = -2\lambda\dot{u}_a, \quad (22)$$

$$\pi_{ab} = 2\lambda\sigma_{ab}. \quad (23)$$

Equation (23) implies $\eta = -\lambda$; that is, negative viscosity, which is physically unacceptable. This case will only be admissible when σ_{ab} vanishes, namely, for a shear-free fluid. In such a case the expansion θ must be positive in order to keep ρ positive.

V. EXAMPLE

Let us now apply the former results to one of the best known vacuum solutions—the Schwarzschild exterior solution.

The Schwarzschild exterior solution,[17]

$$ds^2 = -\left(1-\frac{2M}{r}\right)dt^2 + \left(1-\frac{2M}{r}\right)^{-1} dr^2 + r^2\, d\Omega^2, \tag{24}$$

is the only stationary and spherically symmetric vacuum solution which becomes flat at $r=\infty$. If we perform a conformal transformation, where the function U must depend on the temporal coordinate, the resulting metric will no longer be stationary (nor spherically symmetric if it also depends on any angular coordinate). But, as was previously pointed out in Sec. II, the Killing vector associated with the temporal coordinate in the Schwarzschild solution generates a conformal motion for the solution

$$\pounds_\xi \tilde{g}_{ab} = 2(\pounds_\xi U)\tilde{g}_{ab}, \quad \xi = \frac{\partial}{\partial t}. \tag{25}$$

The velocity field u_a transforms itself under the action of ξ as[24]

$$\pounds_\xi u_a = (\pounds_\xi U)u_a, \quad \pounds_\xi u^a = -(\pounds_\xi U)u^a. \tag{26}$$

Hereafter we shall work under two main assumptions.

(a) The function U depends only on t and r; $U = U(t,r)$.

(b) It is of the form

$$U = (1/\alpha)(f(t) + g(r))^\alpha. \tag{27}$$

The first assumption together with (26) leads to

$$u_{t,t} = U_t u_t, \quad u_{r,t} = U_t u_r, \tag{28}$$

and after some straightforward calculations we come to the conclusion that

$$U_t = C(r)U_r, \quad u_t = C(r)u_r, \tag{29}$$

where $C(r)$ is an undetermined function of the radial coordinate. Here λ will then be

$$\lambda = e^{-U}|U_r|\{(1-2M/r)^{-1}C^2 - (1-2M/r)\}. \tag{30}$$

By introducing (28) and (29) into (27) we obtain

$$U = \frac{K_1}{\alpha}\left(k_2 + t + \int \frac{dr}{C(r)}\right)^\alpha,$$

$$U_r = \frac{K_1}{C(r)}\left(k_2 + t + \int \frac{dr}{C(r)}\right)^{\alpha-1}, \tag{31}$$

$$u_t = C(r)u_r, \quad u_r = (1/\lambda)U_r, \tag{32}$$

$$\dot{u}_t = -\frac{e^{-2U}(1-2M/r)^2 C'}{(1-2M/r)^{-1}C^2 - (1-2M/r)}u_r,$$

$$\dot{u}_r = -\frac{e^{-2U}CC'}{(1-2M/r)^{-1}C^2 - (1-2M/r)}u_r, \tag{33}$$

$$\rho = 3\lambda^2 + 4Q + 2R, \quad p = -\lambda^2 - 2P + \tfrac{2}{3}(R-Q), \tag{34}$$

where P, Q, and R are, respectively,

$$P = e^{-2U}U_{rr}\{(1-2M/r)^{-1}C^2 - (1-2M/r)\}, \tag{35}$$

$$Q = e^{-2U}U_r\{(1-2M/r)^{-1}CC' - 2(r-M)/r^2\}, \tag{36}$$

$$R = e^{-2U}U_r\left\{2\frac{r-M}{r^2} + \frac{M}{r^2}\frac{(1-2M/r)^{-1}C^2 + (1-2M/r) - (r^2/M)CC'}{(1-2M/r)^{-1}C^2 - (1-2M/r)}\right\}, \tag{37}$$

where a prime indicates differentiation with respect to r and U_{rr} stands for the second derivative of the function U with respect to r.

In order to match continuously this solution to the Schwarzschild exterior solution, U_t, U_r, and p must vanish on the matching surface defined by $U(t,r) = 0$. The necessary and sufficient condition for this to occur is $\alpha \geqslant 3$. The matching surface $U(t,r) = 0$, according to (30), is given by

$$k_2 + t + \int \frac{dr}{C(r)} = 0. \tag{38}$$

This equation gives, for every time t, the radius $r_M(t)$ of the collapsing object under consideration; and it must be a decreasing function of the time. This fact implies that for $r < r_M$ the quantity $t + k_2 + \int [dr/C(r)]$ is negative; and it makes K_1 negative in the case of odd α and positive when α is even.

The only restriction upon the function $C(r)$ is that the following energy conditions[23] have to be satisfied:

$$\text{(i)}\quad \rho > 0, \qquad \text{(ii)}\quad -\rho \leqslant p_i \leqslant \rho, \tag{39}$$

where P_i are the spacelike eigenvalues of the stress-energy tensor under consideration. A sufficient (but not necessary) condition for (ii) to hold is

$$-\rho \leqslant 3p \leqslant \rho. \tag{40}$$

There are a number of functions $C(r)$ satisfying these requirements, for instance,

$$C(r) = (2M/na^2)(r/2M + b)^{n+1}, \tag{41}$$

where a and b are constants that must be chosen, together with n, to satisfy the above requirements. Fitting values for such constants [and for the other constants appearing in (31)] are, for example,

$$\alpha = 5,7,9, \quad M = 600, \quad a^2 = 10^{-4}, 10^{-5}, \quad n = 5,$$
$$b = -7.5, \quad K_1 = 10^{-24}, \quad k_2 = -160. \tag{42}$$

The density ρ and pressure p so obtained readily verify conditions (38) and (39) for every time t in the region $2M < r < r_M$, and they are both decreasing functions of the radius r.

The radius $r_M(t)$ of the collapsing object is given by

$$r_M(t) = 2M\{[a^2/(t+K_2)]^{1/n} - b\}, \tag{43}$$

which obviously decreases with time.

The time interval for which all the former results hold varies from $t_0 = 0$ to $t_1 = a^2(1+b)^{-n} - k_2$; at this time (t_1) the fluid has "collapsed" at the gravitational radius $r = 2M$. For $t > t_1$ there is a singularity $(t = k_2)$ and the model is no longer valid.

On the surface $r = 2M$ the metric obviously becomes singular like the pressure and the density. The behavior of the present solution in the region $r < 2M$ deserves a more accurate study, which is being carried out at present, but it is beyond the scope of the present work. Nevertheless, it is interesting to note that for $t > -k_2$ [recall that $t = -k_2$ is a singularity for $r_M(t)$], the radius of the collapsing object becomes greater than $2M$ and it tends to $-2Mb$ as t tends to infinity, which is a finite value greater than $2M$.

Another way of avoiding the singularity on the surface $r = 2M$ could consist of matching a surface $r_0 > 2M$ with another metric, in a similar way to that of Coley and Tupper.[25]

An appropriate choice of $C(r)$ could possibly lead us to a pulsating solution [pulsating $r_M(t)$].

Several additional examples may be found in more simple cases; i.e., when the function U depends only on t, either in the case of the Schwarzschild solution or even in the more simple case of Minkowski's flat space-time, giving in this case an expanding perfect fluid solution and then reproducing some of the plane FRW models.

ACKNOWLEDGMENT

We are grateful to Dr. J. Ibanez for his encouragement and helpful discussions in this work.

[1]L. Herrera and J. Ponce de Leon, J. Math. Phys. **26**, 2847 (1985).
[2]M. Cosenza, L. Herrera, M. Esculpi, and L. Witten, J. Math. Phys. **22**, 118 (1981).
[3]L. Herrera, G. J. Ruggeri, and L. Witten, Astrophys. J. **234**, 1094 (1979).
[4]L. Herrera, J. Jimenez, and G. J. Ruggeri, Phys. Rev. D **22**, 2305 (1980).
[5]M. Cosenza, L. Herrera, M. Esculpi, and L. Witten, Phys. Rev. D **25**, 2527 (1982).
[6]C. Eckart, Phys. Rev. **58**, 919 (1940).
[7]L. Landau and E. M. Lifshitz, *Fluid Mechanics* (Addison–Wesley, Reading, MA, 1958).
[8]A. Lichnerowicz, *Theories Relativistes de la Gravitation et de l'Electromagnetisme* (Masson, Paris, 1955).
[9]G. F. R. Ellis, "Relativistic cosmology," in *General Relativity and Gravitation*, XLII Enrico Fermi Summer School Proceedings, edited by R. K. Sachs (Academic, New York, 1971).
[10]V. Canuto, "Neutron stars. General review," Solvay Conference on Astrophysics and Gravitation, Brussels, 1973.
[11]V. A. Belinskii, E. S. Nikomarov, and S. M. Khalatnikov, Sov. Phys. JETP **50**, 213 (1979).
[12]A. A. Coley and B. O. J. Tupper, Astrophys. J. **271**, 1 (1983).
[13]B. O. J. Tupper, J. Math. Phys. **22**, 2666 (1981).
[14]A. K. Raychaudhuri and S. K. Saha, J. Math. Phys. **23**, 2554 (1982).
[15]J. Carot and J. Ibanez, J. Math. Phys. **26**, 2282 (1985).
[16]A. Z. Petrov, *Einstein Spaces* (Pergamon, Oxford, 1969).
[17]D. Kramer, H. Stephani, M. Mc. Callum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Springer, Berlin, 1980).
[18]Ch. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[19]There exist some generalizations to Eckart's theory,[6] due to Israel[20,21] and Hiscock and Lindblom.[22] For further details, see this last reference.
[20]W. Israel, Ann. Phys. (NY) **100**, 310 (1976).
[21]W. Israel and J. M. Stewart, Ann. Phys. (NY) **118**, 341 (1979).
[22]W. A. Hiscock and L. Lindblom, Phys. Rev. D **31**, 725 (1985).
[23]S. W. Hawking and G. F. R. Ellis, *The Large Structure of Space-Time* (Cambridge U.P., Cambridge, 1973).
[24]K. Yano, *The Theory of Lie Derivatives and its Applications* (North-Holland, Amsterdam, 1957).
[25]A. A. Coley and B. O. J. Tupper, Phys. Rev. D **29**, 2701 (1984).

A class of perfect fluid metrics with flat three-dimensional hypersurfaces

Thomas Wolf
Sektion Physik, Friedrich Schiller Universität Jena, Max-Wien-Platz 1, GDR 6900 Jena, German Democratic Republic

(Received 19 March 1986; accepted for publication 30 April 1986)

The class of perfect fluid and vacuum space-times with a family of flat three-slices and a tensor of exterior curvature covariantly constant within these slices is examined and the corresponding solutions are found. It is shown that this class contains the class of metrics with three commuting Killing vectors. Therefore, e.g., all known stationary metrics with cylindrical or plane symmetry are generalized. An instruction is given for constructing perfect fluid metrics with this symmetry and a connection to a vacuum across surfaces $p = 0$. Thereby the equation of state of the interior rotating perfect fluid can be arbitrarily chosen and the positivity of density and pressure can be forced. A geometric criterion of the interior metric with rotating matter is found that decides whether the exterior solution is stationary or static. Besides solutions with three symmetries, inhomogeneous metrics also are found. Among them is a solution with one symmetry and rotating, expanding, shearing, and accelerating perfect fluid. All resulting vacuum solutions are already known.

I. INTRODUCTION AND FORMALISM

With this paper a further step shall be undertaken in the program of examining metrics with flat three-dimensional slices. The proposal to assume the existence of hypersurfaces with special inner symmetries was first pointed out by Collins and Szafron. Thereby an approach to inhomogeneous metrics was suggested without supposing space-time symmetries or the Weyl tensor being algebraically special from the very beginning. On the other hand, at the end of the third of their papers[1] they doubted the derivation of new solutions with the strongest such restriction, with flat slices, because first attempts failed even with the help of a computer. Nevertheless in Refs. 2–4 the property of flat slices was established for many already known solutions, and new solutions with this property also were found. The vital point for obtaining these results consisted of allowing the flat slices to be timelike, not to impose restrictions on the four-velocity u^α but to impose algebraic conditions on the tensor of the exterior curvature $K_{\alpha\beta}$.

Because the projection K_{ab}, $a,b = 1,2,3$, of this tensor is a three-tensor within the flat hypersurfaces $x^4 = \text{const}$ and the hypersurfaces are invariantly defined because of their flatness, restrictions on K_{ab} possess a geometric meaning and are to be preferred against restrictions on metric components. In Ref. 2 perfect fluid solutions with vanishing K_{ab} were investigated. Despite zero inner and exterior curvature the metrics contain a Petrov type D solution describing rotating dust without symmetry. In Ref. 3 the assumption $K_{ab} = \lambda g_{ab}$ leads to new perfect fluid solutions with three Killing vectors. In Ref. 4 vacuum solutions of the Petrov types I, III, and N, and partly without symmetries, result under the condition $K^a{}_a = 0$. Thereby the Petrov type correlates to the rank of K_{ab}: for rank = 1 only type N solutions result, for rank = 2 besides pp-waves of type N a type III solution of Kundt's class[5,6] is contained, and for rank = 3 a special solution is found, the Petrov[7] solution of type I.

In this paper the demand on K_{ab} takes differential character. We assume that K_{ab} is covariantly constant within the flat hypersurfaces, i.e., $K_{ab\|c} = 0$. A consequence will be that at least one Killing vector exists; on the other hand, classes of solutions with Petrov type I also result. The advantage of the assumption $K_{ab\|c} = 0$ is to be mathematically helpful but not to be very restrictive. It provides the $(4,a)$-components of the field equations to be equivalent to the demand that for timelike flat slices the four-velocity u^α lies within these slices, i.e., $u^4 = 0$, and that for spacelike flat slices u^α is orthogonal to them, i.e., $u_a = 0$. In the timelike case the components u_a are free for being defined by the (a,a)-components of the field equations satisfied that way. The conventions and the formalism are already given in Ref. 3 and shall therefore be depicted in a concentrated form.

The family of flat hypersurfaces is parametrized by the timelike (resp. spacelike) coordinate x^4. We denote the normal vector by $n_\alpha = (0,0,0,N)$, $n^\alpha n_\alpha = \epsilon = \pm 1$ and the flat three-metric by $g_{ab}(x^4)$. With the three-vector N^a, which fixes the mutual position of coordinate systems of neighboring slices, the space-time metric reads

$$ds^2 = g_{ab}(dx^a + N^a dx^4)(dx^b + N^b dx^4) + \epsilon(N\,dx^4)^2. \tag{1.1}$$

Here $g^{\alpha\beta}$ and n^α take, with the inverse three-metric $\overset{3}{g}{}^{ab}$, the form

$$g^{\alpha\beta} = \begin{pmatrix} \overset{3}{g}{}^{ab} + \epsilon N^a N^b/N^2 & -\epsilon N^a/N^2 \\ -\epsilon N^b/N^2 & \epsilon/N^2 \end{pmatrix}, \tag{1.2}$$

$$n^\alpha = (-N^a/N, 1/N).$$

Latin indices are moved by g_{ab} and $\overset{3}{g}{}^{ab}$. With the projection tensor

$$h_{\alpha\beta} = g_{\alpha\beta} - \epsilon n_\alpha n_\beta,$$

the tensor of exterior curvature $K_{\alpha\beta}$ is defined to be

$$K_{\alpha\beta} = -n_{\alpha;\gamma}h^{\gamma}{}_{\beta}.$$

The normalization and vanishing of rotation of n_α yield

$$K_{\alpha\beta} = K_{\beta\alpha}, \quad K_{\alpha\beta}n^\beta = 0,$$

$$\rightarrow K_{4a} = N^b K_{ba}, \quad K_{44} = N^a N^b K_{ab}.$$

The three-part K_{ab} reads, with the covariant three-derivative $_{\|a}$,

$$K_{ab} = (N_{a\|b} + N_{b\|a} - g_{ab,4})/2N. \tag{1.3}$$

As described in Refs. 8 (§21) and 9 the space-time curvature tensor can be given in a comprehensive form by K_{ab}, n_α, their covariant derivatives, and $\underset{n}{\pounds} K_{ab}$. Formulated in the antisymmetric part of $N_{a,b}$ and the symmetric part corresponding to K_{ab}, the Ricci tensor is

$$\epsilon N R_{ab} = g_{r(b}K^r{}_{a),4} - K_{ab\|r}N^r - K_{ab}K^r{}_r N + K^r{}_b N_{[a,r]} + K^r{}_a N_{[b,r]} - N_{,a\|b}, \tag{1.4a}$$

$$\epsilon N R^4{}_m = K^b{}_{b\|m} - K^b{}_{m\|b}, \tag{1.4b}$$

$$\epsilon(R - 2R_{ab}\overset{3}{g}{}^{ab}) = K^a{}_a{}^2 - K^a{}_b K^b{}_a. \tag{1.4c}$$

In Sec. II an approach for solving the field equations with a perfect fluid is developed. The cases corresponding to the four possible algebraic types of $K^m{}_n$ are investigated in Secs. III–VI. In the summary a table containing all subcases is given.

II. THE STRUCTURE OF THE FIELD EQUATIONS AND THE APPROACH FOR THEIR SOLUTION

With $K_{ab\|c} = 0$ the field equations take the form

$$(\mu + p)u^4 u_m = 0, \tag{2.1}$$

$$\epsilon(K^a{}_a{}^2 - K^a{}_b K^b{}_a) = 2(\mu + p)u_n u_b \overset{3}{g}{}^{nb} - 2\mu, \tag{2.2a}$$

$$\epsilon(K^a{}_a{}^2 - K^a{}_b K^b{}_a) = 2p + 2(\mu + p)\epsilon N^2(u^4)^2, \tag{2.2b}$$

$$g_{r(b}K^r{}_{a),4} - K_{ab}K^r{}_r N + K^r{}_b N_{[a,r]} + K^r{}_a N_{[b,r]} - \epsilon N_{,a\|b} = \epsilon N(\mu + p)u_a u_b + \epsilon N g_{ab}(\mu - p)/2. \tag{2.3}$$

In (2.2b) the norm of the four-velocity

$$-1 = u_\alpha u_\beta g^{\alpha\beta} = u_a u_b \overset{3}{g}{}^{ab} + \epsilon N^2(u^4)^2 \tag{2.4}$$

was used.

The three equations (2.1) are no longer differential equations, but for $\mu \neq p$ follows the vanishing of $u^4 u_b$. Because u^α is timelike the norm gives

$$\text{for } \epsilon = -1, \quad u_a = 0, \quad 1 = N^2(u^4)^2, \tag{2.5a}$$

$$\text{for } \epsilon = 1, \quad u^4 = 0, \quad -1 = u_a u_b \overset{3}{g}{}^{ab}. \tag{2.5b}$$

In addition to an appropriate equation of state, (1.3) are to be satisfied, which are called the K_{ab}-equations in the following, because K_{ab} is no longer defined by them but N_a and N are determined with given K_{ab}. In most cases two of the field equations will be regarded as defining μ and p and afterwards what equation of state they satisfy will be examined.

The first step for solving the field equations will be to choose appropriate coordinates in the three-surface. Because the inner curvature of the three-surface vanishes we could assume $g_{ab} = \mathrm{diag}(-\epsilon,1,1)$ with $K_{ab\|c} = 0$ resulting in $K_{ab}(x^4)$. If we now want to align the coordinate axis to the eigenvectors of K_{ab} we must perform x^4-dependent coordinate transformations that make the three-metric x^4-dependent. We therefore assume from the beginning

$$g_{ab} = g_{ab}(x^4).$$

With this choice we can replace the covariant derivative $_\|$ by the partial derivative and we have $K_{ab,c} = 0$, i.e.,

$$K^a{}_b = K^a{}_b(x^4).$$

To align the coordinate axis to the eigenvectors of K_{ab} we need a survey of the possible normal forms of a three-dimensional symmetric tensor. In a spacelike three-surface g_{ab} is positive definite and K_{ab} takes diagonal form. Concerning the timelike case it is known from the literature[10,11] that in four dimensions a symmetric tensor defines invariantly a two-subspace of the tangent space by transforming every vector of the two-surface into a vector of the same two-surface. The following four different cases occur.

In case A, the two-surface is timelike and contains A1, two real orthogonal eigenvectors; A2, no real eigenvector; or A3, a double real null eigenvector. In case B, the two-surface is null and contains a threefold null eigenvector.

The corresponding normal forms of K_{mn} are

A1: $K_{mn} = \lambda_1 x_m x_n + \lambda_2 y_m y_n + \lambda_3 z_m z_n$,

$g_{mn} = -\epsilon x_m x_n + y_m y_n + z_m z_n, \epsilon = 1$,

A2: $K_{mn} = \lambda_1 k_{(m}l_{n)} + \lambda_2(l_m l_n - k_m k_n) + \lambda_3 z_m z_n$,

A3: $K_{mn} = \lambda_1 k_{(m}l_{n)} + \lambda_2 k_m k_n + \lambda_3 z_m z_n$,

B: $K_{mn} = \lambda_1(z_m z_n - 2k_{(m}l_{n)}) + \lambda_2 l_{(m}z_{n)}$,

where g_{ab} is given in the cases A2, A3, and B by

$$g_{mn} = z_m z_n - 2k_{(m}l_{n)}.$$

Nonvanishing scalar products of the vectors x^n, y^n, z^n, k^n, and l^n are $x^n x_n = y^n y_n = z^n z_n = 1$, $k^n l_n = -1$. The eigenvalues (resp. combinations of them) will be assigned later with $a(x^4)$, $b(x^4)$, and $c(x^4)$. The case of spacelike three-surfaces is contained in A1 for $\epsilon = -1$.

In order to solve the field equations we make use of some of them for defining kinematic quantities. Because $u^4 u_m = 0$ from (2.1), Eq (2.2) defines μ (resp. p). A combination of μ and p is obtained from the trace of (2.3), making use of the norm $u_a u_b \overset{3}{g}{}^{ab} = -1$, which is valid for timelike hypersurfaces. By that $\mu - p$ can be replaced in (2.3).

Summarizing, the solution contains in every one of the four cases A1–B (i) the choice of coordinates in correspondence with the algebraic structure of K_{ab} and thereby the determination of the components of $K^a{}_b(x^4)$ and $g_{ab}(x^4)$; (ii) the derivation of N, N_a from the six linear partial differential equations of first order

$$2N(x^\alpha)K_{ab}(x^4) = 2N_{(a}(x^\alpha)_{,b)} - g_{ab}(x^4); \tag{2.6}$$

(iii) the solution of the field equations

$$g_{r(b}K^r{}_{a),4} - K_{ab}K^r{}_rN + K^r{}_bN_{[a,r]} + K^r{}_aN_{[b,r]}$$

$$= \begin{cases} -N_{,ab} + g_{ab}(K^r{}_{r,4} - K^r{}_r{}^2N + N^{,r}{}_{,r})/3, \\ \qquad\qquad \text{for } \epsilon = -1, \\ N_{,ab} + g_{ab}(K^r{}_{r,4} - K^r{}_mK^m{}_rN - N^{,r}{}_{,r}) \\ + N(\mu+p)u_au_b, \quad \text{for } \epsilon = 1. \end{cases} \tag{2.7}$$

(iv) the calculation of μ and p from

$$2\mu = K^a{}_a{}^2 - K^a{}_bK^b{}_a,$$
$$6p = (4K^a{}_{a,4} - NK^a{}_a{}^2 - 3NK^a{}_bK^b{}_a + 4N^{,a}{}_{,a})/N,$$
$$\text{for } \epsilon = -1, \tag{2.8a}$$
$$2p = K^a{}_a{}^2 - K^a{}_bK^b{}_a,$$
$$2\mu = (4K^a{}_{a,4} + NK^a{}_a{}^2 - 5NK^a{}_bK^b{}_a - 4N^{,a}{}_{,a})/N,$$
$$\text{for } \epsilon = 1; \tag{2.8b}$$

(v) for $\epsilon = -1$ the determination of u_a from (2.7) and of u_4 from $u_4 = N_a u_b \overset{3}{g}{}^{ab}$.

In the following sections (III–VI) the cases A1–B will be treated this way.

III. SOLUTIONS WITH K_{mn} OF TYPE A1

Type A1 implies that two real eigenvectors are contained in the two-surface invariantly defined by K_{mn}. Following the above-described approach, we obtain, with the choice of coordinates

$$x_m = x_{,m}, \quad y_m = y_{,m}, \quad z_m = z_{,m},$$

for $K^m{}_n$ and g_{mn},

$$K^m{}_n = \text{diag}(a(x^4), b(x^4), c(x^4)),$$
$$g_{mn} = \text{diag}(-\epsilon, 1, 1). \tag{3.1}$$

From three of the six equations $NK_{mn} = N_{(m,n)} = 0$, for $m \neq n$, we get, e.g.,

$$(NK_{12})_{,3} - (NK_{31})_{,2} + (NK_{23})_{,1} = N_{2,13} = 0,$$

the existence of functions $D_m(x^\alpha)$, so that

$$N_1 = (D_3 - D_2)_{,1}, \quad N_2 = (D_1 - D_3)_{,2},$$
$$N_3 = (D_2 - D_1)_{,3}, \tag{3.2}$$
$$D_{m,m} = 0 \quad \text{(without summation)}.$$

The analysis of the three residual K_{mm}-equations and that of all further steps depend on $\text{rank}(K_{mn})$ and the value $\epsilon = \pm 1$.

A. Rank(K_{mn}) = 3, spacelike slices

Before examining the field equations, the three remaining K_{mm}-equations are to be solved. Performing appropriate differentiations and comparing coefficients without specifying ϵ, we obtain from (2.6), (3.1), and (3.2) longer expressions for N_m that are quadratic in x^a and an expression for N that is linear in x^a. The coefficients are products of seven arbitrary x^4-dependent functions and the x^4-dependent $K^m{}_n$ components a,b,c. From the field equations (2.7) we obtain linear expressions in x^a for $N(\mu+p)u_mu_n$.

Specifying $\epsilon = -1$ and assuming *differing eigenvalues of $K^a{}_b$ by pairs*, i.e., the algebraic type in Segré notation [111], with $u_m = 0$ from (2.5a) the vanishing of six of the seven x^4-dependent functions follows and from this $N_a \sim x^a$. Such a metric is transformable by $\bar{x}^a = q_1^a(x^4)x^a + q_2^a(x^4)$ (without summation) into a form $g_{mn}(x^4)$, $N_m = 0$, while conserving $N_{,m} = 0$ so that three Killing vectors occur. Because these solutions with a G_3I on S_3 are examined several times and are known for special equations of state (Ref. 12, §12.4.), they are not further investigated.

Assuming $\epsilon = -1$ and $a = b \neq c$, i.e., the algebraic type [(11)1], with $u_m = 0$ the vanishing of five of the seven functions follows. Because we still have for $K^1{}_1 = K^2{}_2$ the freedom of a x^1,x^2-rotation, also the sixth function can be chosen equal to zero. In analogy to the above case the solution has the Bianchi type I and is not further investigated.

If we finally have equal eigenvalues of K_{mn} and therefore $K_{mn} \sim g_{mn}$, i.e., the type [(111)], then we are restricted to a subcase of metrics, already treated in Ref. 3, that belong to a class of conformally flat space-times found by Stephani.[13]

B. Rank(K_{mn}) = 3, timelike slices

By having the K_{ab}-equations totally solved at the beginning of the above section, we have to expose the ten free x^4-dependent functions (seven by integration, three by $K^m{}_n$) to the remaining field equations

$$(N(\mu+p)u_au_b)^2 = N(\mu+p)u_a{}^2N(\mu+p)u_b{}^2.$$

Inserting the linear-in-x^a expressions for $N(\mu+p)u_a{}^2$ and comparing coefficients of x^a we see that three of the seven functions must be zero. One function can be transformed to 1 by a transformation $\bar{x}^4(x^4)$ with the consequence $N = 1$. The comparison of x^a-coefficients for determining the three remaining functions depends on the equality of the $K^m{}_n$-eigenvalues.

At first, *differing eigenvalues by pairs* are assumed, i.e., the type [11,1]. Performing the residual comparisons of x^a-coefficients we obtain the following metric:

$$\epsilon = 1, \tag{3.3a}$$
$$g_{ab} = \text{diag}(-1,1,1), \tag{3.3b}$$
$$N = 1,$$
$$N_1 = -ax^1 + hx^2 - sx^3, \tag{3.3c}$$
$$N_2 = -hx^1 + bx^2 + gx^3, \tag{3.3d}$$
$$N_3 = sx^1 - gx^2 + cx^3, \tag{3.3e}$$

$$h = [(b+c)_{,4} - b^2 - c^2 + a(b+c)]^{1/2}$$
$$\times [-(a+c)_{,4} + a^2 + c^2 - b(a+c)]^{1/2}/(b-a),$$
$$g = [-(c+a)_{,4} + c^2 + a^2 - b(c+a)]^{1/2}$$
$$\times [-(b+a)_{,4} + b^2 + a^2 - c(b+a)]^{1/2}/(c-b),$$
$$s = [-(a+b)_{,4} + a^2 + b^2 - c(a+b)]^{1/2}$$
$$\times [(c+b)_{,4} - c^2 - b^2 + a(c+b)]^{1/2}/(a-c),$$

where $a(x^4)$, $b(x^4)$, and $c(x^4)$ are possibly only restricted by an equation of state $\mu(p)$ with

$p = ab + bc + ca,$

$\mu = 2(a+b+c)_{,4} + ab + bc + ca - 2(a^2 + b^2 + c^2),$

and the inequalities

$$u_1{}^2 = ((b+c)_{,4} - b^2 - c^2 + a(b+c))/(\mu + p)$$
$$(= 1 + u_2{}^2 + u_3{}^2 > 0),$$
$$u_2{}^2 = (-(c+a)_{,4} + c^2 + a^2 - b(c+a))/(\mu + p)$$
$$(\geq 0),$$
$$u_3{}^2 = (-(a+b)_{,4} + a^2 + b^2 - c(a+b))/(\mu + p)$$
$$(\geq 0).$$

We can choose the functions a,b,c, and their x^4-derivatives in a slice $x^4 = x_0^4$ so that these relations are fulfilled at least in the neighborhood of a slice $x^4 = \text{const}$. Generalizing the regularity demand of this section, solutions with $\operatorname{rank}(K_{ab}) = 2$, i.e., without loss of generality $c = 0$, are contained in (3.3). After a lengthy calculation the Killing equations yield with

$$\xi^\alpha T^\beta{}_{\beta,\alpha} = 0 = \xi^\alpha (T^\gamma{}_\beta T^\beta{}_\gamma)_{,\alpha} \to \xi^\alpha{}_{\mu,\alpha} = 0 = \xi^4,$$
$$\xi^a = \xi^a(x^4), \quad \xi^a{}_{,4} = A^a{}_b \xi^b,$$
$$A^a{}_b = \begin{pmatrix} -a & h & -s \\ h & -b & -g \\ -s & g & -c \end{pmatrix},$$

i.e., according to theorems on the existence of solutions for systems of first-order ordinary differential equations the metric has a $G_3 I$ on T_3. Because, in general, $\mu = \text{const} \cdot p$ does not hold, no further homothetic vector exists. The four-velocity is nondiverging and nongeodesic and has shear and rotation. Setting μ and p equal to zero, the resulting vacuum solution is the Kasner metric with the Petrov type I, which therefore also occurs in the case of perfect fluid. Because a vacuum metric of Das[14] fulfills the assumptions of this section (III B) with the necessary result of the Kasner solution, both metrics could be identified by coordinate transformations.[15] The further contained already known solutions with perfect fluid will be given in Sec. IV A in context with other very similar new solutions.

If *two of the three $K^a{}_b$-eigenvalues coincide*, then the cases $a = b \neq c$, i.e., $[1(1,1)]$, and $a \neq b = c$, i.e., $[(11),1]$, occur. Comparing x^a-coefficients in the remaining identities,

$$(u_a u_b)^2 = u_a{}^2 u_b{}^2,$$

as described at the beginning of this section (III B), we obtain $u_1 = 0$ for $a = b$, in contrast to timelike u^α and $u_2 = u_3 = 0$ and the vanishing of two further functions for $b = c$. With $K^2{}_2 = K^3{}_3$ and a therefore possible x^2,x^3-rotation also the third function can be set to zero. We get for the metric

$$\epsilon = 1, \quad g_{ab} = \operatorname{diag}(-1,1,1),$$
$$N = 1, \quad N_1 = -ax^1, \quad N_2 = bx^2, \quad N_3 = bx^3. \quad (3.4)$$

Here $a(x^4)$, $b(x^4)$ are restricted by

$$0 = (a+b)_{,4} - a(a-b) \quad (3.5)$$

and possibly an equation of state with

$$p = 2ab + b^2, \quad \mu = 2b_{,4} - 3b^2. \quad (3.6)$$

If the equation of state is not to be specified, a and b can be given with arbitrary $k(x^4)$ by $(k-b)(k-2b) - k_{,4} = 0$ and $a = k - b$. The four-velocity $u_\alpha = (1,0,0,-ax^1)$ is nondiverging, nonrotating, nonshearing, and nongeodesic. From the Killing equations a lengthy calculation yields, with $\xi^\alpha \mu_{,\alpha} = 0 = \xi^4$ and $0 = \underset{\xi}{\pounds}\, u^\alpha$ from $0 = u_\alpha \underset{\xi}{\pounds}\, T^{\alpha\beta}$, the existence of a G_4 on T_3. Because u_α shows acceleration but no rotation, the metric belongs to the LRS metrics, class II.[16,17] Setting μ and p equal to zero, the Kasner metric of the "pancake" type with timelike flat slices results.

The residual case of *equal eigenvalues of K_{mn}* and therefore $K_{mn} \sim g_{mn}$, i.e., the type $[(11,1)]$, is a subcase of the metrics already treated in Ref. 3 and leads to the de Sitter universe.

C. Rank(K_{mn}) = 2

By assuming diagonal form of K_{ab}, which characterizes Sec. III, and further assuming that $K^3{}_3 = 0$, we attribute a special role to x^3. Thus we have to consider the cases $g_{mn} = \operatorname{diag}(\pm 1, 1, \mp 1)$ and $g_{mn} = \operatorname{diag}(1,1,1)$. Having solved with (3.2) three of the six K_{mn}-equations, the remaining equations with $K_{11} = ag_{11}$, $K_{22} = b$, $K_{33} = 0$ shall be treated for the three cases above. The equations read

$$N = (D_3 - D_2)_{,11}/g_{11}a \quad (3.7a)$$
$$= (D_1 - D_3)_{,22}/b, \quad (3.7b)$$
$$0 = (D_2 - D_1)_{,33}. \quad (3.7c)$$

With $D_{a,a} = 0$ [because of (3.2)] we can integrate (3.4c) by introducing functions of two variables. Comparing x^3-coefficients, we obtain, with functions $l(x^4), s(x^4), w(x^4)$, and $H(x^1,x^2,x^4)$,

$$N_1 = H_{,1} - x^3 x^1 L/b - x^3 s, \quad (3.8a)$$
$$N_2 = -Lx^3x^2/(g_{11}a) - H_{,2} + wx^3, \quad (3.8b)$$
$$N_3 = sx^1 - wx^2 + Lx^{1^2}/(2b) + Lx^{2^2}/(2ag_{11}), \quad (3.8c)$$
$$N = (bH_{,11} - x^3 L)/(abg_{11}), \quad (3.8d)$$

and for H the equation

$$0 = bH_{,11} + g_{11}aH_{,22}. \quad (3.9)$$

By that three of the field equations (2.7) take the form

$$\epsilon N(\mu + p)u_1u_2 = H_{,12}(b - a) - \epsilon H_{,1112}/(ag_{11}), \quad (3.10a)$$
$$\epsilon N(\mu + p)u_2u_3 = (lx^2/(ag_{11}) - w)b, \quad (3.10b)$$
$$\epsilon N(\mu + p)u_3u_1 = (lx^2/b + s)a. \quad (3.10c)$$

Because of $N_{,33} = K^3{}_3 = 0$ the 3,3-component of the field equation (2.7) reads

$$(\mu + p)u_3{}^2 = -g_{33}(\mu - p)/2. \quad (3.11)$$

Starting from (3.11), assertions concerning the equation of state will be made later. From the K_{ab}-equations only (3.9) is still to be solved.

In the case of *spacelike slices* ($\epsilon = -1, g_{11} = 1$), we are restricted to metrics with a $G_3 I$ on S_3, as an investigation of (3.10) and the three other field equations (2.7) implies. By the implications that H is quadratic in x^1 and x^2 it follows that $N_{,a}=0$ and $N_a \sim x^a$ and, in analogy to Sec. III A three commuting Killing vectors are obtained. This case has been treated several times (Ref. 12, §12.4) and is not further investigated.

For *both remaining cases of timelike slices*

$$g_{ab}=\operatorname{diag}(g_{11},1,-g_{11}),\quad g_{11}=\pm 1,$$

at first general assertions are made before they are handled separately.

Computing $N^2(\mu+p)^2u_2{}^2u_3{}^2$ from the 2,3-component and on the other hand from the 2,2- and 3,3-component of the field equations (3.10) and (2.7) and comparing x^3-coefficients, we obtain $L=0$. In the following we assume

$$N_{,a}\neq 0, \tag{3.12}$$

because $N_{,a}=0$ implies that H is quadratic in x^1,x^2, which corresponds to the cases $c=0$ (resp. $a=0$) of solution (3.3). At first the residual K_{ab}- equation (3.9) will be solved by introducing new variables y,z and functions $f(y,x^4),g(z,x^4)$, which are conjugate complex for $abg_{11}>0$:

$$y=x^1+\sqrt{(-b/ag_{11})}x^2, \tag{3.13a}$$

$$z=x^1-\sqrt{(-b/ag_{11})}x^2, \tag{3.13b}$$

$$H=(f(y,x^4)+g(z,x^4))ag_{11}. \tag{3.14}$$

Derivatives with respect to y and z are designated with primes in the following. The identities

$$(N(\mu+p)u_au_b)^2=N(\mu+p)u_a{}^2N(\mu+p)u_b{}^2,$$

with $N(\mu+p)u_au_b$ from (2.7) and (3.14), remain to be satisfied. With the field equation (3.11) we can draw conclusions for the equation of state:

$$g_{11}=-1\rightarrow(\mu+p)(\mu-p)<0\quad\text{or}\quad u_3=\mu-p=0, \tag{3.15a}$$

$$\begin{aligned} g_{11}=1\rightarrow u_3{}^2&=1+u_1{}^2+u_2{}^2\\ &\rightarrow(\mu+3p)(\mu+p)<0\\ &\quad\text{or}\quad \mu+3p=u_1=u_2=0. \end{aligned}\tag{3.15b}$$

Only for $g_{11}=-1$, $u_3=\mu-p=0$ will the solutions be given and the method obtained. Hints are given for the unphysical cases.

Under the assumptions

$$u_3=\mu-p=0,\quad g_{ab}=\operatorname{diag}(-1,1,1), \tag{3.16}$$

we can solve the $3,a$-components of the field equations (2.7) and (3.10) easily by $w=s=0$ and by separation of $f(y,x^4)$ and $g(z,x^4)$. Integrating both resulting ordinary differential equations of second order we obtain for f'' and g'' exponential expressions in y (resp. z). Inserting these in the last identity $(u_1u_2)^2=u_1{}^2u_2{}^2$, $f'''g'''=0$ follows, without loss of generality, $g'''=0$ (because of the symmetry $y\leftrightarrow z$, $x^2\leftrightarrow -x^2$). In order not to derive the metric (3.3) we agreed upon in (3.12), $0=N_{,2}=(b/a)f'''$ ($\in\mathbb{R}$), i.e., with (2.8b), we have

$$\mu=p=ab>0.$$

For $p_{,4}=0$ and $p_{,4}\neq 0$ the remaining identity $(u_1u_2)^2=u_1{}^2u_2{}^2$ provides different metrics.

For $p_{,4}=(ab)_{,4}=0$ we obtain with the replacements $b=n^2/a$, $n=\text{const}$, and $F=f'$, $G=g'$, the metric

$$\epsilon=1, \tag{3.17a}$$

$$g_{ab}=\operatorname{diag}(-1,1,1), \tag{3.17b}$$

$$N=F_{,1}+G_{,1}, \tag{3.17c}$$

$$N_1=-a(F+G), \tag{3.17d}$$

$$N_2=n(F-G), \tag{3.17e}$$

$$N_3=0, \tag{3.17f}$$

$$\begin{aligned} F=&-\left(\frac{n^2-a^2}{n^4+a^4}\right)^{1/2}\\ &\times n_1\cos\left[\left(\frac{n^4+a^4}{n^2-a^2}\right)^{1/2}\left(x^1+\frac{n}{a}x^2\right)+n_2\right]\\ &+a_{,4}\,\frac{(a^4+n^4-4a^2n^2)}{2(a^2-n^2)(a^4+n^4)}\left(x^1+\frac{n}{a}x^2\right), \end{aligned}\tag{3.17g}$$

$$G=a_{,4}\,\frac{(x^1-x^2n/a)}{2(a^2-n^2)}, \tag{3.17h}$$

$$\mu=p=n^2,\quad a^2<n^2,$$

$$u_\alpha=(n,a,0,2anF)(n^2-a^2)^{-1/2}.$$

Here $n_1(x^4)$, $n_2(x^4)$, $n=\text{const}\neq 0$ can be chosen freely.

By a $\bar{x}^4(x^4)$-transformation $a=x^4$ can be reached for $a_{,4}\neq 0$ and $n_1=1$ for $a_{,4}=0$.

Properties of the metric: The four-velocity is nondiverging and geodesic and has rotation but no shear. The derivation of all symmetries is a lengthy calculation despite the relations $0=\underset{\xi}{\pounds}\mu=\underset{\xi}{\pounds}u_\alpha$, which result from $0=\underset{\xi}{\pounds}T^\alpha{}_\alpha=\underset{\xi}{\pounds}T^{\alpha\beta}T_{\alpha\beta}=u_\alpha\,\underset{\xi}{\pounds}T^{\alpha\beta}$. For $a_{,4}\neq 0$ the Killing vectors form a G_2I on T_2:

$$\xi^\alpha=(n,-a,0,0)/(n^2-a^2)^{1/2},$$

$$\xi^\alpha=(0,0,1,0).$$

For $a_{,4}=0$ we have five Killing vectors

$$\xi^\alpha=(0,0,1,0),\quad \xi^\alpha=(n,-a,0,0),$$

$$\begin{aligned}\xi^\alpha=(&q_{,4}a(n^2-a^2)/\sin\varphi-2n^4q,\\ &q_{,4}n(n^2-a^2)/\sin\varphi+2a^3nq,0,\\ &(n^2-a^2)^{1/2}(n^4+a^4)^{1/2}\\ &\times(q_{,4}\cot\varphi+(2n^2q-q_{,44})/n_{2,4})),\end{aligned}$$

with

$$\varphi=((n^4+a^4)/(n^2-a^2))^{1/2}(x^1+(n/a)x^2)+n_2$$

and q satisfying

$$0=q_{,4}n_{2,4}+((q_{,44}-2n^2q)/n_{2,4})_{,4}.$$

An additional homothetic vector does not exist. The solution has the Petrov type D and setting μ and p equal to zero, the space-time is flat. Because of the special equation of state $\mu=p=\text{const}$, no surfaces $p=0$ exist for connecting the vacuum. Because u_α has rotation, the solution cannot be generated out of vacuum by a procedure of Tabensky and Taub[18] and Wainwright.[19] For $a=x^4$ the solution seems not to be of high physical interest because it has high symmetry and the projections of u_α, $R_{\alpha\beta}$ to normalized vectors as $v^\alpha=\delta^\alpha_1$ become singular and therefore true singularities occur.

For $a=\text{const}$, a theorem of Ozsváth[20] and Farnsworth and Kerr[21] shows that the Gödel solution[22] is the only homogeneous perfect fluid solution with a maximal G_5 [see also the discussion of the properties of the metrics (4.8) and (4.9)].

In the case in Sec. III C, $p_{,4} = (ab)_{,4} \neq 0$ we can integrate the remaining identity $(u_1u_2)^2 = u_1{}^2u_2{}^2$ with respect to x^4. With the replacements $F = f'$, $G = g'$ we obtain the metric

$$\epsilon = 1, \tag{3.18a}$$

$$g_{ab} = \mathrm{diag}(-1,1,1), \tag{3.18b}$$

$$N = F_{,1} + G_{,1}, \tag{3.18c}$$

$$N = -a(F+G), \tag{3.18d}$$

$$N = (ab)^{1/2}(F-G), \tag{3.18e}$$

$$N = 0, \tag{3.18f}$$

$$F = \left(x^1 + \left(\frac{b}{a}\right)^{1/2} x^2\right) \frac{(a-b)_{,4}(a+b)}{2(a-b)(a^2+b^2)} + \sum_{\pm} n_{\pm} \left(\frac{a-b}{(a^2+b^2)a}\right)^{1/2} \times \exp\left[\pm \left(\frac{(a^2+b^2)a}{a-b}\right)^{1/2}\left(x^1 + \left(\frac{b}{a}\right)^{1/2} x^2\right)\right], \tag{3.18g}$$

$$G = \frac{(a_{,4}(a-3b) - b_{,4}(b-3a))}{2(a^2+b^2)(a-b)} \left(x^1 - \left(\frac{b}{a}\right)^{1/2} x^2\right), \tag{3.18h}$$

$$\mu = p = ab \tag{3.18i}$$

$$u_1{}^2 = (ab)_{,4}/(2n(a+b)N) + b/(b-a), \tag{3.18j}$$

$$u_2{}^2 = (ab)_{,4}/(2n(a+b)N) + a/(b-a), \tag{3.18k}$$

$$u_3 = 0, \tag{3.18l}$$

$$u_4 = -u_1N_1 + u_2N_2. \tag{3.18m}$$

Here $n_+(x^4), n_-(x^4)$ can be chosen freely and they are conjugate complex if $b > a$, if not they are real. Also $a(x^4)$ and $b(x^4)$ satisfy

$$0 = ab(a-b) - n(a+b),\ 0 = n = \text{const}, \quad ab > 0, \quad (ab)_{,4} \neq 0. \tag{3.19}$$

Properties of the metric: The four-velocity is not geodesic and it diverges and rotates and has shear. The calculation provides the only Killing vector ∂_3 and no further homothetic vector. Setting μ and p equal to zero the Minkowski space-time results. The vacuum cannot be connected across surfaces $p = 0$ because the components of $u_\alpha, R_{\alpha\beta}$ and also the projections with normalized vectors as $v^\alpha = \delta_1^\alpha$ become singular for $0 = a + b$, which results from (3.19) and $p = ab = 0$. On the contrary this metric is remarkable as a cosmological solution and especially in the mathematical respect. According to an analysis of Wainwright[23,24] and as far as the author knows, it is the first spatial inhomogeneous metric with rotating and diverging perfect fluid. But just such metrics of high generality are necessary to answering cosmological questions.[1,25,26] Both metrics (3.17) and (3.18) have the properties that u^α lies in the plane spanned by the repeated principal null vectors l^α, k^α of the Weyl tensor, i.e., $u^{[\alpha}k^\beta l^{\gamma]} = 0$, and the magnetic part of the Weyl tensor with respect to u^α vanishes. According to Carminati and Wainwright[27] the first property applies to almost all known type D metrics except the solutions of Wahlquist[28] and Kramer.[29] With the metrics (3.17) and (3.18), examples are given for a theorem of Carminati and Wainwright[27] that Petrov type D solutions with an equation of state $p = p(\mu)$, the property $u^{[\alpha}k^\beta l^{\gamma]} = 0$, and the vanishing magnetic part of the Weyl tensor with respect to u^α have one of the following properties: the equation of state satisfies $dp/d\mu = 0$ or 1, or at least three Killing vectors exist. For the metrics (3.17) and (3.18), $p = \mu$ is valid.

Having treated above the only physically sensible case (3.16) of the distinction (3.15), some hints for solving the other cases will be given. To achieve linear equations in f and g and perform separations, we take advantage of the fact that the field equations (3.10b) and (3.10c) contain neither f nor g and that from the field equations (2.7) follows $N(\mu + p)(g_{11}au_1{}^2 + bu_2{}^2) = -(ab)_{,4}$. Linear equations can be obtained by replacing u_a with the field equations (2.7) and (3.10) in the identities

$$0 = N^2(\mu+p)^2(u_1{}^2(u_2u_3)^2 - u_2{}^2(u_1u_3)^2),$$

$$0 = N^2(\mu+p)^2(g_{11}a(u_1u_3)^2 + b(u_2u_3)^2) + N(\mu+p)u_3{}^2(ab)_{,4}.$$

Because positivity conditions are violated, the metrics are not further investigated.

D. Rank(K_{mn}) = 1

With (3.2) we have solved three of the K_{mn}-equations for $m = n$. For the remaining three we assume $K^1{}_1 = a$, $K^2{}_2 = K^3{}_3 = 0$. With the result $0 = N_{,22} = N_{,23} = N_{,33}$ the 2,2-, 2,3-, and 3,3-components of the field equation (2.7) provide $0 = u_2 = u_3 = \mu - p$ and (2.8) provides in addition $\mu = p = 0$. For a vacuum Minkowski space-time follows immediately from the expressions for the curvature tensor.

The above treatment includes all cases of type A1, i.e., diagonalizable K_{mn}. Two types of metrics result. One group has rank(K_{mn}) = 3, the Petrov type I, a free equation of state, and at least as G_3I on T_3, whereas the other metrics have rank(K_{mn}) = 2, the Petrov type D, the equation of state $\mu = p$, and sometimes only one Killing vector.

The following section treats the case of tensors K_{mn} of type A2. Because they can be diagonalized by complex transformations we will become acquainted with two groups of metrics with the same algebraic properties as in this section.

IV. SOLUTIONS WITH K_{mn} OF TYPE A2

According to the approach described in Sec. II, we will again choose appropriate coordinates and components of $K^m{}_n$ and g_{mn} to solve the K_{ab}-equations and then the field equations (2.7). The emphasis lies on simplifying $K^m{}_n$ as far as possible, i.e., that $K^m{}_n$ takes second Jordan's normal form according to the algebraic type A2 (see Ref. 30, p. 175):

$$K^m{}_n = \begin{pmatrix} 0 & -a^2 & 0 \\ 1 & 2b & 0 \\ 0 & 0 & c \end{pmatrix}, \tag{4.1a}$$

$$K_{mn}=\begin{pmatrix}1 & 0 & 0\\ 0 & -a^2 & 0\\ 0 & 0 & c\end{pmatrix}, \tag{4.1b}$$

$$g_{mn}=\begin{pmatrix}2b/a^2 & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}, \tag{4.1c}$$

$$\overset{3}{g}{}^{mn}=\begin{pmatrix}0 & 1 & 0\\ 1 & -2b/a^2 & 0\\ 0 & 0 & 1\end{pmatrix}, \tag{4.1d}$$

$$a(x^4),b(x^4),c(x^4),a^2>b^2.$$

The (in the following not further applied) eigenvalues λ_i and eigenvectors k^n,l^n,z^n in

$$K_{mn}=\lambda_1 k_{(m}l_{n)}+\lambda_2(l_m l_n-k_m k_n)+\lambda_3 z_m z_n$$

and

$$g_{mn}=z_m z_n-2k_{(m}l_{n)}$$

take the form

$$k^m=-(a^2-b^2)^{1/4}\begin{pmatrix}0\\ 1/a\\ 0\end{pmatrix},$$

$$l^m=(a^2-b^2)^{-1/4}\begin{pmatrix}a\\ -b/a\\ 0\end{pmatrix},\quad z^m=\begin{pmatrix}0\\ 0\\ 1\end{pmatrix},$$

$$\lambda_1=-2b,\quad \lambda_2=-(a^2-b^2)^{1/2},\quad \lambda_3=c.$$

Because of the diagonal form of K_{mn} and $g_{mn,4}$, from the three K_{mn}-equations (2.6) for $m\neq n$ in analogy to (3.2), the existence of functions D_a with $D_{a,a}=0$ (without summation) follows, as do

$$N_1=(D_3-D_2)_{,1},\quad N_2=(D_1-D_3)_{,2},$$
$$N_3=(D_2-D_1)_{,3}. \tag{4.2}$$

The analysis of the remaining three K_{mm}-equations differs for $K_{33}=c=0$ and $c\neq 0$.

A. Rank(K_{mn}) = 3

For regular K_{mn} of the type $[z\bar{z},1]$ the analysis of the remaining three K_{mm}-equations yields results analogous to those in case A1 in Secs. III A and III B. The expressions for N_a and N are quadratic (resp. linear) in x^n. Again seven free x^4-dependent functions occur that, together with $a(x^4),b(x^4),c(x^4)$ from $K^m{}_n$, have to satisfy

$$(N(\mu+p)u_a u_b)^2=N(\mu+p)u_a{}^2 N(\mu+p)u_b{}^2.$$

Replacing for $N(\mu+p)u_m u_n$ the (in x^n) linear expressions from the field equations (2.7), it results that three of the seven functions vanish. A futher function can be set equal to 1 by a $\bar{x}^4(x^4)$-transformation and the three residual functions are connected with a,b,c by three algebraic equations.

1. The metric

$$\epsilon=1, \tag{4.3a}$$

$$g_{mn}=\begin{pmatrix}2b/a^2 & 1 & 0\\ 1 & 0 & 1\\ 0 & 0 & 1\end{pmatrix}, \tag{4.3b}$$

$$N=1 \tag{4.3c}$$

$$N_1=(1+(b/a^2)_{,4})x^1+gx^2-sx^3, \tag{4.3d}$$

$$N_2=-gx^1-a^2x^2+wx^3, \tag{4.3e}$$

$$N_3=sx^1-wx^2+cx^3. \tag{4.3f}$$

A possible approach to fulfilling the three remaining quadratic equations for real functions a,b,c,g,s,w of x^4 and to guaranteeing the positivity conditions $u_a{}^2\geq 0$, $\mu\geq p\geq 0$, consists of the following steps.

(i) Choosing a function $K^a{}_a(x^4)$, so that for a given equation of state $\mu=\mu(p)$ the relation $\mu\geq p\geq 0$ is satisfied by

$$(\mu-p)/2=K^a{}_{a,4}-K^a{}_a{}^2+2p\geq 0. \tag{4.4a}$$

(ii) Choosing $b(x^4)$, so that a real function v can be defined and u_3 can be computed by means of the relations

$$0\leq v^2=p+3b^2-2bK^a{}_a, \tag{4.4b}$$

$$0\leq(\mu+p)u_3{}^2/2=-b_{,4}+bK^a{}_a-p. \tag{4.4c}$$

(iii) Defining a function q by

$$4v^2q^4=(v_{,4}-vK^a{}_a)^2 + ((\mu+p)u_3{}^2/2+p)((\mu+p)u_3{}^2/2+\mu). \tag{4.4d}$$

(iv) Determining a from

$$a^2=v^2+b^2. \tag{4.4e}$$

(v) For being able to compute u_2 from

$$0<(\mu+p)u_2{}^2/a^2=2q^2-v_{,4}/v+K^a{}_a, \tag{4.4f}$$

substituting if necessary $x^4\to -x^4$, $b\to -b$, $K^a{}_a\to -K^a{}_a$, and satisfying thereby also the above demands.

(vi) Computing g from

$$g=q^2-v_{,4}/(2v)+a_{,4}/a. \tag{4.4g}$$

(vii) Computing u_1 from

$$(\mu+p)u_1{}^2=-b(\mu-p)/a^2-K^a{}_a+2g$$
$$[=(\mu+p)(u_1u_2)^2/u_1{}^2>0]. \tag{4.4h}$$

(viii) Computing c,s,w from

$$c=K^a{}_a-2b, \tag{4.4i}$$

$$N(\mu+p)u_2u_3=(c-2b)s-a^2w, \tag{4.4j}$$

$$N(\mu+p)u_3u_1=-cs-w, \tag{4.4k}$$

with a unique solution because the coefficient determinant of s,w does not vanish with $a^2>b^2$.

2. Properties of the metric

The four-velocity is not geodesic, is nondiverging, and has shear and rotation. From the Killing equations it follows that, after a lengthy calculation with $\xi^\alpha\mu_{,\alpha}=0=\xi^4$,

$$\xi^n = \xi^n(x^4), \quad \xi^n{}_{,4} = A^n{}_m \xi^m,$$

$$A^n{}_m = \begin{pmatrix} g & a^2 & -w \\ -1-(b/a^2)_{,4} - 2gb/a^2 & -g-2b & s+2wb/a^2 \\ -s & w & -c \end{pmatrix},$$

i.e., according to theorems on the existence of first-order systems of ordinary differential equations the metric has a $G_3 I$ on T_3. Because μ and p do not, in general, satisfy $\mu = \text{const} \cdot p$, no further homothetic vector exists. As described below the specialization to vacuum by setting μ and p equal to zero has the Petrov type I, which is the most general one and is therefore also valid for a perfect fluid.

3. Subdivision of the metric into an interior and an exterior solution

An important question for investigating space-times with perfect fluid concerns the possibility for connecting a vacuum solution. In the case of the metric (4.3) this can be answered easily. Because the spatial coordinate x^4 parametrizes the flat slices and, on the other hand, $\mu(x^4)$ and $p(x^4)$ can be chosen freely, it is possible to describe perfect fluid and a vacuum and their connection by one single metric. To be of physical interest, the solution has to guarantee that $\mu \geq p \geq 0$ in the whole interior. Moreover the equation of state should not be restricted to extreme cases ($p = 0, p = \mu$). It is shown in Appendix B that these demands can be fulfilled for the metric (4.3). For this purpose it is necessary to choose functions $K^a{}_a(x^4)$ and $b(x^4)$ in such a way that, for a given equation of state, the inequalities (4.4a)–(4.4c) can be satisfied in the interval $0 < x^4 \leq x_0^4$ and that at x_0^4 the functions have the values $K^a{}_a = -1/x_0^4$, $b = b_0/x_0^4$, and $b_0 = \text{const}$, for a connecting vacuum.

In order to interpret x^4 as a radial coordinate in such a metric, the integral curves of an appropriate combination of both spatial Killing vectors have to satisfy the periodicity condition, i.e., the metric must be constituted in such a way that points (t,r,φ,z) and $(t,r,\varphi+\pi,z)$ can be identified. Moreover space-time should be regular on the axis of symmetry. In this case the quotient of the circumference and the radius of small circles gives 2π.

Summarizing, it can be stated that the metric (4.3) describes the interior and exterior space of a stationary nondiverging perfect fluid with rotation, shear, a free choice of the equation of state, and $\mu \geq p \geq 0$ throughout, if appropriate functions $K^a{}_a(x^4)$ and $b(x^4)$ are determined (e.g., see Appendix B) and x^4 can be interpreted as a radial coordinate or the metric is symmetric to $x^4 = 0$.

4. The vacuum case

The vacuum solution corresponding to (4.3) reads, for $K^a{}_a = 0$,

$$g_{mn} = \begin{pmatrix} 1/(2b) & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{4.5a}$$

$$\varepsilon = 1, \tag{4.5b}$$

$$N_1 = x^1, \tag{4.5c}$$

$$N_2 = -4b^2 x^2, \tag{4.5d}$$

$$N_3 = -2bx^3, \tag{4.5e}$$

$$N = 1, \quad b = \text{const.} \tag{4.5f}$$

For positive b this is identical to the solution

$$k^2 ds^2 = dx^2 + e^{-2x} dy^2 + e^x [\cos(\sqrt{3}x)(dz^2 - dt^2) - \sin(\sqrt{3}x) dz\, dt]$$

of Petrov[7] and for negative b identical with it after a transformation $\bar{z} = iz, \bar{t} = it$.

For $K^a{}_a \neq 0$ the vacuum solution reads

$$g_{mn} = \begin{pmatrix} x^4/(1-2m) & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{4.6a}$$

$$\epsilon = 1, \tag{4.6b}$$

$$N_1 = x^1(1 + 1/(2(1-2m))) - x^2/2x^4, \quad N_2 = x^1/2x^4 + 2m(1-2m)x^2/(x^4)^2,$$

$$N_3 = (2m-1)x^3/x^4, \quad N = 1, \quad m = \text{const}, \quad m \neq \tfrac{1}{2}, \quad 0 \leq 3m^2 - 2m.$$

The limiting case $m = \frac{2}{3}$ corresponds to $K^m{}_n$ of type A3 (see Sec. V). The metric has the Petrov type I and a $G_3 I$ on T_3. For $m = \frac{2}{3}$ the Killing vectors are

$$\xi = \partial_3, \quad \xi = 2(x^4)^{1/6}\partial_1 + 3(x^4)^{7/6}\partial_2, \quad \xi = (2\ln(x^4) - 3)(x^4)^{1/6}\partial_1 + 3(x^4)^{7/6}\ln(x^4)\partial_2,$$

and for $m \neq \frac{2}{3}$ they read, with the abbreviations $M = (3m^2 - 2m)^{1/2}$, $\varphi = M\ln(x^4)$,

$$\xi = (x^4)^{1-2m}\partial_3, \quad \xi = (x^4)^{m-1/2}(m\cos\varphi + M\sin\varphi)\partial_1 + (x^4)^{m+1/2}\cos\varphi\,\partial_2,$$

$$\xi = (x^4)^{m-1/2}(m\sin\varphi + M\cos\varphi)\partial_1 + (x^4)^{m+1/2}\sin\varphi\,\partial_2.$$

After a first transformation in coordinates $\bar{x}^a$ so that the Killing vectors read $\partial_{\bar{a}}$, a second transformation $\bar{x}^1 = i\hat{x}^1 + i\hat{x}^2$, $\bar{x}^2 = -\hat{x}^1 + \hat{x}^2$ provides a metric in diagonal form:

$$ds^2 = (x^4)^{2m}\frac{(3m^2-2m)^{1/2}}{2m-1}[-(d\hat{x}^1)^2(x^4)^{2i(3m^2-2m)^{1/2}}(i(1-m) + (3m^2-2m)^{1/2})$$
$$+ (d\hat{x}^2)^2(x^4)^{-2i(3m^2-2m)^{1/2}}(i(1-m) - (3m^2-2m)^{1/2})] + (x^4)^{2-4m}(d\bar{x}^3)^2 + (dx^4)^2.$$

For $a^2 < b^2$, i.e., $3m^2 - 2m < 0$, it is the Kasner metric with $K^m{}_n$ of type A1. This is in accordance with the possibility of transforming tensors $K^m{}_n$ of the form (4.1) with $a^2 < b^2$ into diagonal form, i.e., these tensors with $a^2 < b^2$ are of type A1 for which the Kasner metric necessarily resulted whereas A2 is characterized by $a^2 > b^2$.

The method of complexifying a metric in order to obtain a real one by a following complex coordinate transformation is called a "complex trick" and leads, e.g., to the derivation of the Kerr–Newman metric starting with the Reissner–Nordström metric. Because mostly stationary metrics were obtained from static ones with this trick, the question occurs if a similar relation also exists for the metrics (3.3) and (4.3). Though the four-velocity shows rotation in both cases, only Killing vectors of the vacuum metric (4.6), which can be connected with the perfect fluid metric (4.3), rotate also, whereas the Killing vectors of the Kasner metric, which can be connected with the metric (3.3), are nonrotating. Because both three-metrics of the surfaces $p = 0$ between perfect fluid and vacuum are flat, the different algebraic types of $K^m{}_n$ are responsible for the vacuum metrics to be static (resp. stationary).

5. Classification of known contained solutions

In Secs. III B and IV A solutions were considered with timelike flat slices, rank$(K_{ab}) = 3$, and $K_{ab\|c} = 0$, and the metrics (3.3), (3.4), (4.3), (4.5), and (4.6) were derived, which admit at least an Abelian G_3 on T_3. On the other hand, all solutions with a G_3I have flat slices, namely the subspaces spanned by the Killing vectors, e.g., in the form ∂_m with consequently $g_{ab,m} = 0$. In addition these slices obey $K_{ab\|m} = 0$ because of $0 = N_{a,b} = N_{,b}$. The investigations of this paper therefore contain all metrics with a G_3I and perfect fluid or vacuum. The metrics given above represent the general case of metrics with a G_3I and perfect fluid or vacuum, whereby the tensor of exterior curvature of the Killing orbit has rank 3. The vacuum solution (4.6) is contained in the class of Lewis[31] as the case of stationary cylindrical symmetry. The solutions (3.3) and (4.3) with perfect fluid are new in this generality with all three free functions of one variable or two free functions and free equation of state. Contained in them are all stationary plane symmetric as well as stationary cylindrically symmetric metrics. A prominent representative of this class is a metric of Krasinski[32] describing a perfect fluid with stationary cylindrical symmetry, rigid rotation, and a vorticity vector parallel to the axis of symmetry. Due to the last two conditions still one free function of one variable remains, the equation of state results from it, and a linear ordinary differential equation of second order has to be solved. This metric can be classed with (3.3) or (4.3) only locally because the type of K_{mn} changes depending on x^4.

As far as the author knows, the metric (4.3) is the first one describing the connection of a rotating cylindrically symmetric perfect fluid across surfaces $p = 0$ with a vacuum and satisfying all positivity conditions in the interior.

B. Rank$(K_{mn}) = 2$

For solving this case we have to start with K_{mn} and g_{mn} in the form (4.1) with $c = 0$ and the solutions (4.2) for N_a from three of the K_{ab}-equations (2.6). In total analogy to Sec. III C the three remaining K_{aa}-equations can be solved with intermediate results corresponding to (3.7)–(3.9) and (3.12)–(3.14). The resulting expressions for N_a and N contain functions f and g of the new variables

$$y = x^1 + ax^2 \quad (\text{resp. } t = x^1 - ax^2).$$

Derivatives with respect to these variables are assigned by primes and overdots. Again the 3,3-component of the field equation (2.7),

$$(\mu + p)u_3{}^2 = -(\mu - p)/2,$$

restricts the equation of state and provides two possibilities:

$$u_3 = \mu - p = 0, \qquad (4.7\text{a})$$

$$u_3 \neq 0, \quad (\mu + p)(\mu - p) < 0. \qquad (4.7\text{b})$$

The treatment of these cases is the same as in Sec. III C. For the only physically meaningful case (4.7a) the difference concerns conditions for a and b. From $N_{,a} \neq 0 < u_1{}^2$, we get $a_{,4} = 0$ and $a > b$, and with the A2 condition, $a^2 > b^2$, further $a > b > -a$.

The special case $a^2 = 2b^2$ must be listed extra, because f and g are restricted by $f'''' = \ddot{g} = 0$ under this assumption. After an appropriate $\bar{x}^4(x^4)$-transformation the metric reads, in coordinates t,z:

$$\begin{aligned} ds &= ((2 \pm \sqrt{2})dy^2 \pm 2\sqrt{2}\,dy\,dt - (2 \mp \sqrt{2})dt^2)/(4a) \\ &\quad + (dx^3)^2 + y(y+m)dt\,dx^4 + ((y+m/2)^2 \\ &\quad - (a/4)y^2(y+m)^2(2 \pm \sqrt{2}))(dx^4)^2, \end{aligned} \qquad (4.8)$$

$$\mu = p = a^2, \quad 0 = u^y = u^3 = u^4, \quad u^t = 2(a/(2 \mp \sqrt{2}))^{1/2}.$$

The function $m(x^4)$ and the constant a, $a>0$, can be chosen freely. A lengthy calculation applying $\underset{\xi}{\pounds} u^\alpha=0$ from $u_\alpha \underset{\xi}{\pounds} T^{\alpha\beta}=0$ shows the existence of five Killing vectors but of no further homothetic vector. They are given by

$$\xi^\alpha=(\xi^t,\xi^y,\xi^3,\xi^4)=(1,0,0,0),\quad \xi^\alpha=(0,0,1,0),$$

$$\xi^\alpha=(2q+(1\pm\sqrt{2})\xi^y,$$

$$q_{,4}(y+m^2/(4y+2m))/2+q_{,4}m/4+m_{,4}q/2$$

$$\mp q_{,44}/(\sqrt{2}a),0,\pm 2q_{,4}/(\sqrt{2}a(2y+m))-q).$$

q obeys

$$0=2q_{,444}/a-aq_{,4}m^2-amm_{,4}q$$

$$+2(2\mp\sqrt{2})q_{,4}\mp\sqrt{2}qm_{,44}\mp 2\sqrt{2}q_{,4}m_{,4}.$$

The metric has the Petrov type D and u_α is nondiverging and geodesic and has rotation but no shear.

With the assumption $a^2\neq 2b^2$ and the substitutions $F=f'$, $G=g'$ the metric reads

$$g_{ab}=\begin{pmatrix} 2b/a^2 & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}, \tag{4.9a}$$

$$\epsilon=1, \tag{4.9b}$$

$$N=F_{,1}+G_{,1}-b_{,4}/(2a^2), \tag{4.9c}$$

$$N_1=F+G+b_{,4}x^1/(2a^2), \tag{4.9d}$$

$$N_2=a(-F+G+b_{,4}x^2/(2a^2)), \tag{4.9e}$$

$$N_3=0, \tag{4.9f}$$

$$F=\left(\frac{a-b}{a^2-2b^2}\right)^{1/2}\sum_{\pm} n_\pm \times\exp\pm\left[\left(\frac{a^2-2b^2}{a-b}\right)^{1/2}(x^1+ax^2)\right] - b_{,4}\frac{(2b^3-4ab^2+3ba^2-2a^3)(x^1+ax^2)}{4a^2(b-a)(a^2-2b^2)}, \tag{4.9g}$$

$$G=b_{,4}b\frac{x^1-ax^2}{4a^2(b-a)},\quad a=\text{const},$$

$$a>b>-a,\quad 0\neq a^2-2b^2, \tag{4.9h}$$

$$\mu=p=a^2, \tag{4.9i}$$

$$u_1=\frac{2b-a}{a(2(a-b))^{1/2}}, \tag{4.9j}$$

$$u_2=\frac{a}{(2(a-b))^{1/2}}, \tag{4.9k}$$

$$u_3=0, \tag{4.9l}$$

$$u_4=\frac{2aF+b_{,4}(x^1-ax^2)/a}{(2(a-b))^{1/2}}. \tag{4.9m}$$

Here $n_+(x^4)$, $n_-(x^4)$ can be chosen freely and are conjugate complex for $a^2<2b^2$ otherwise real. With a $\bar{x}^4(x^4)$-transformation it is possible to obtain $b=x^4$ for $b_{,4}\neq 0$ (resp. for constant b):

$$|n_+n_-|=1 \quad\text{if}\quad n_+n_-\neq 0 \quad\text{or}\quad |n_\pm|=1 \quad\text{if}\quad n_\pm\neq 0.$$

The Killing equations are solved with $\underset{\xi}{\pounds} u^\alpha=0$ from $u_\alpha \underset{\xi}{\pounds} T^{\alpha\beta}=0$. For $b_{,4}\neq 0$ the Killing vectors read $\xi^\alpha=(0,0,1,0)$ and $\xi^\alpha=(a,-1,0,0)/(b-a)^{1/2}$. They form a G_2I on T_2. For $b_{,4}=0$ and $|n_+n_-|=1$ the metric has five Killing vectors. With the abbreviations S,D for the sum and the difference of

$$n_+\exp+\left[\left(\frac{a^2-2b^2}{a-b}\right)^{1/2}(x^1+ax^2)\right]$$

and

$$n_-\exp-\left[\ \right],$$

they read

$$\xi^\alpha=(0,0,1,0),\quad \xi^\alpha=(a,-1,0,0),$$

$$\xi^\alpha=\Big(a(a-b)q_{,4}/D+a^2(a-2b)q,$$

$$-(2b+a)(a-b)q_{,4}/(aD)+2a^2q,$$

$$0,(a-b)^{1/2}(a^2-2b^2)^{1/2}$$

$$\times\left[\frac{q_{,4}S}{4D}-\frac{2n_+}{n_{+,4}}\left(a^2q+\frac{q_{,44}}{8}\right)\right]\Big),$$

q satisfies

$$0=q_{,4}n_{+,4}/n_+\ [(n_+/n_{+,4})(8qa^2+q_{,44})]_{,4}.$$

A further homothetic vector does not exist. The metric has the Petrov type D and u_α is nondiverging and geodesic and has rotation but no shear.

Properties of the metrics (4.8) and (4.9): For both metrics, the magnetic part of the Weyl tensor vanishes with respect to u^α and the principal null directions of the Weyl tensor l^α,k^α satisfy $u^{[\alpha}k^{\beta}l^{\gamma]}=0$. With the solutions (4.8) and (4.9) [as previously with (3.17) and (3.18)], examples are given for a theorem of Carminati and Wainwright[27] stating that Petrov type D metrics, with an equation of state $p=p(\mu)$, the property $u^{[\alpha}k^{\beta}l^{\gamma]}=0$, and the vanishing magnetic part of the Weyl tensor with respect to u^α, have one of the following properties: the equation of state satisfies $dp/d\mu=0$ or 1 or at least three Killing vectors exist. For the metrics (4.8) and (4.9), $\mu=p$ is valid.

The metric (4.9) with $b=x^4$ has the disadvantage that projections of u_α and $R_{\alpha\beta}$ to normalized vectors as $v^\alpha=(a,0,0,0)/\sqrt{|2b|}$ become singular for $x^4\to a$ and therefore true singularities occur.

The solutions (4.9) with $b=\text{const}$, (3.17) with $a=\text{const}$, and (4.8) are homogeneous space-times with perfect fluid and a maximal G_5. A theorem of Ozsvath[20] and Farnsworth and Kerr[21] states that the Gödel solution[22]

$$ds^2=\bar{a}^2(d\bar{x}^2+d\bar{y}^2+e^{2\bar{x}}d\bar{z}^2/2-(d\bar{t}+e^{\bar{x}}d\bar{z})^2)$$

(here written with an overbar to prevent confusion) is the only one with perfect fluid and a maximal G_5. In one direction the transformation can be given easily. The Gödel solution takes the form of (4.9) with

$$b=0,\quad a=(\sqrt{2}\bar{a})^{-1},\quad n_+=\bar{a}/\sqrt{2},\quad n_-=0,$$
$$x^1=\sqrt{\bar{a}}(\bar{x}-\bar{t})/\sqrt[4]{8},\quad x^2=\sqrt{\bar{a}^3}(\bar{t}+\bar{x})/\sqrt[4]{2},$$
$$x^3=\bar{y}/(\sqrt{2}a),\quad x^4=\bar{z}.$$

Furthermore it can be shown easily that the Gödel solution has a one-parameter manifold of families of flat slices: for every function $x(\bar{x},\bar{z})$ the slices $x=$ const are flat [the manifold is one-parametric because any new function $\hat{x}=\hat{x}(x)$ gives the same family of slices].

With the derivation of the metrics (4.8) and (4.9), the first of the two cases (4.7) of this section (IV B) is treated. The unphysical second case could be treated in analogy to the corresponding case for K_{ab} of type A1 at the end of Sec. III C.

Because the type A2 does not allow $\operatorname{rank}(K_{ab})=1$, this section is finished.

V. SOLUTIONS WITH K_{mn} OF TYPE A3

As described in Sec. II appropriate coordinates are chosen so that $K^m{}_n$ takes the second Jordan's normal form, which corresponds to the algebraic type

$$K^m{}_n=\begin{pmatrix}0 & -a^2 & 0\\ 1 & 2a & 0\\ 0 & 0 & c\end{pmatrix}, \tag{5.1a}$$

$$K_{mn}=\begin{pmatrix}1 & 0 & 0\\ 0 & -a^2 & 0\\ 0 & 0 & c\end{pmatrix}, \tag{5.1b}$$

$$g_{mn}=\begin{pmatrix}2/a & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}, \tag{5.1c}$$

$$\overset{3}{g}{}^{mn}=\begin{pmatrix}0 & 1 & 0\\ 1 & -2/a & 0\\ 0 & 0 & 1\end{pmatrix}, \tag{5.1d}$$

$$a=a(x^4),\quad c=c(x^4).$$

The eigenvalues and eigenvectors in

$$K_{mn}=\lambda_1 k_{(m}l_{n)}+\lambda_2 k_m k_n+\lambda_3 z_m z_n,$$
$$g_{mn}=z_m z_n-2k_{(m}l_{n)}$$

take the form

$$k^m=\begin{pmatrix}-a\\ 1\\ 0\end{pmatrix},\quad l^m=\begin{pmatrix}0\\ 1/a\\ 0\end{pmatrix},\quad z^m=\begin{pmatrix}0\\ 0\\ 1\end{pmatrix},$$
$$\lambda_1=-2a,\quad \lambda_2=-1,\quad \lambda_3=c.$$

A comparison with K_{mn} and g_{mn} from (4.1) shows that (5.1) is the limit of (4.1) for $a=b$ [or $a=-b$ if a is replaced by $-a$ in (5.1)]. The derivations of Secs. IV A and IV B can be applied almost completely.

In the case $\operatorname{rank}(K_{mn})=3$, the solutions (4.3) with $a^2=b^2$ and (4.6) with $m=\frac{2}{3}$ result.

In Sec. IV the symmetry $a\to -a$ existed for $\operatorname{rank}(K_{mn})=2$. Therefore the condition $f'''g'''=0$ for $f(x^1+ax^2)$ and $g(x^1-ax^2)$ could be decided by $g'''=0$ without loss of generality. To take advantage of the calculations of Sec. IV with the decision $g'''=0$, both limits $a=\pm b$ must be investigated. In the fully elaborated case (4.7a) only $a=-b$ is compatible with $f'''=0$. Different solutions result for $a=\pm b$ in the unphysical case (4.7b).

Section V is now finished.

VI. SOLUTIONS WITH K_{mn} OF TYPE B

The approach will be outlined only because it will prove that neither vacuum metrics nor metrics with perfect fluid exist. We have

$$K^m{}_n=\begin{pmatrix}0 & 0 & c^3\\ 1 & 0 & -3c^2\\ 0 & 1 & 3c\end{pmatrix},\quad K_{mn}=\begin{pmatrix}c & 0 & 0\\ 0 & 0 & c^4\\ 0 & c^4 & 3c^5\end{pmatrix},$$

$$g_{mn}=\begin{pmatrix}3 & c & 0\\ c & 0 & 0\\ 0 & 0 & c^4\end{pmatrix},$$

$$\overset{3}{g}{}^{mn}=c^{-4}\begin{pmatrix}0 & c^3 & 0\\ c^3 & -3c^2 & 0\\ 0 & 0 & 1\end{pmatrix},\quad c=c(x^4).$$

The eigenvalues and eigenvectors in

$$K_{mn}=\lambda_1(z_m z_n-2k_{(m}l_{n)})+\lambda_2 l_{(m}z_{n)},$$
$$g_{mn}=z_m z_n-2l_{(m}k_{n)},$$

take the form

$$k^a=\begin{pmatrix}0\\ -1/c^2\\ 0\end{pmatrix},\quad l^a=\frac{1}{c}\begin{pmatrix}c^2\\ -2c\\ 1\end{pmatrix},\quad z=\frac{1}{c^2}\begin{pmatrix}0\\ -c\\ 1\end{pmatrix},$$
$$\lambda_1=c,\quad \lambda_2=2.$$

The solution of the K_{mn}-equations proves to take much more effort than in the other cases because K_{mn} does not have diagonal form. By introducing potentials, by integration and thereby introducing functions of less variables and enabling comparisons of coefficients, the overdetermined system of the six K_{mn}-equations for the four functions N_a, N can be solved successively. Thereby expressions result for N_a and N that are quadratic (resp. linear) in x^a as in Secs. III B and IV A. The remaining equations $(N(\mu+p)u_a u_b)^2=N^2(\mu+p)^2u_a{}^2u_b{}^2$ with expressions for $N(\mu+p)u_a u_b$ from (2.7) provide the vanishing of the seven x^4-dependent free functions. After a $\bar{x}^4(x^4)$-transformation to get $N=1$ the residual equations read

$$2c^3u_1u_2=-c^4+3c^2u_2{}^2-u_3{}^2, \tag{6.1a}$$
$$2c^2u_3u_1=-4c^4-3c^4u_1{}^2+8c^2u_2{}^2-3u_3{}^2, \tag{6.1b}$$
$$2cu_2u_3=-c^4-c^4u_1{}^2+3c^2u_2{}^2. \tag{6.1c}$$

Because $2c^3u_2u_1-2cu_2u_3=c^4u_1{}^2-u_3{}^2$, the two cases $c^2u_1-u_3=0$ and $\neq 0$ result with (6.1b) in $c=0$ or $c^2<0$. Therefore no solutions with perfect fluid and K_{mn} of type B exist. Also, vacuum solutions are excluded by $p=6c^2=0$ from (2.8b).

VII. SUMMARY

The subject of the investigations has been metrics with perfect fluid and a vacuum with a family of flat slices $x^4 = \text{const}$, whose tensor of exterior curvature K_{ab} is covariantly constant within these hypersurfaces. To simplify calculations, coordinates were chosen so that $K^a{}_b$ simplifies as far as possible by taking the second Jordan's normal form. In Table I an overview of the resulting cases is given. Besides the listed metrics with perfect fluid the following vacuum

TABLE I. Overview of the metrics with perfect fluid and flat slices with $K_{ab\|c} = 0$.

Type of K_{mn}	Character of the slices	Segré notation	Properties of the metric
A1, rank 3	spacelike	[111], [(11)1]	G_3I on S_3, known for special equations of state, not further investigated
A1, rank 3	spacelike	[(111)]	special case of Stephani,[a] is rederived in Stephani and Wolf[b]
A1, rank 3	timelike	[11,1]	(3.3), new, G_3I on T_3, Petrov type I, u_α nongeodesic, nondiverging, with shear and rotation
A1, rank 3	timelike	[1(1,1)]	$\nexists$
A1, rank 3	timelike	[(11),1]	(3.4), contained in LRS II
A1, rank 3	timelike	[(11,1)]	de Sitter metric, is rederived in Stephani and Wolf[b]
A1, rank 2	spacelike	[11]	G_3I on S_3, known for special equations of state, not further investigated
A1, rank 2	timelike	[1,1]	(3.17), G_2 on T_2, Petrov type D, u_α nondiverging, geodesic, without shear, with rotation, $\mu = p$
A1, rank 2	timelike	[1,1]	(3.18), new, G_1 on S_1, Petrov type D, u_α not geodesic, expanding, with shear and rotation, $\mu = p$
A1, rank 2	timelike	[1,1]	$(\mu + p)(\mu - p) < 0$
A1, rank 2	timelike	[11]	$(\mu + 3p)(\mu + p) \leq 0$
A1, rank 1			$\nexists$
A2 (timelike)	rank 3	[1,$z\bar{z}$]	(4.3), new, G_3I on T_3, Petrov type I, u_α nondiverging, nongeodesic, with shear and rotation
A2 (timelike)	rank 2	[$z\bar{z}$]	(4.8) Gödel solution
A2 (timelike)	rank 2	[$z\bar{z}$]	(4.9), G_2 on T_2, Petrov type D, u_α nondiverging, geodesic, without shear, with rotation, $\mu = p$
A2 (timelike)	rank 2	[$z\bar{z}$]	$(\mu + p)(\mu - p) < 0$
A3 (timelike)		[1,2]	limits of the metrics of A2 for $4(K^1{}_2)^2 = (K^2{}_2)^2$
B (timelike)		[3]	$\nexists$

[a] H. Stephani, Commun. Math. Phys. **4**, 137 (1967).
[b] H. Stephani and T. Wolf, "Perfect fluid and vacuum solutions of Einstein's field equations with flat 3-dimensional slices," in *Axisymmetric Systems, Galaxies and Relativity: Essays Presented to W. B. Bonnor on His 65th Birthday* (Cambridge U.P., Cambridge, 1985).

solutions are contained: for $K^m{}_n$ with the Segré types [111], [11,1], [(11),1] the Kasner metric, and for $K^m{}_n$ with the Segré type $[1,z\bar{z}]$ with $K^a{}_a = 0$ a metric of Petrov[6] and with $K^a{}_a \neq 0$ the vacuum metrics with stationary cylindrical symmetry. Because a solution of Das[14] satisfies the assumptions of Sec. III it could be identified to be the Kasner metric.[33]

The resulting metrics with perfect fluid can be divided essentially into two groups with the following characteristic properties: (i) $\operatorname{rank}(K_{mn}) = 3$, Petrov type I, a four-velocity without expansion, a free equation of state, and three commuting Killing vectors; or (ii) $\operatorname{rank}(K_{mn}) = 2$, Petrov type D, the equation of state $\mu = p$, and sometimes only one Killing vector. As shown at the end of Sec. IV A, all metrics with three commuting Killing vectors possess flat slices with $K_{mn\|a} = 0$ and therefore are contained in these two groups. Among metrics with stationary plane symmetry is that also contained as a special case: a metric of Krasiński[32] for describing perfect fluid with stationary cylindrical symmetry, rigid rotation, and a vorticity vector parallel to the axis of symmetry. Whereas in this solution one function of one variable can be chosen freely and a second-order ordinary differential equation has to be solved, in (4.3) three functions of one variable or two functions and the equation of state can be chosen freely, so that differential rotation is described also.

The metric (4.3) is, as far as the author knows, the first one that describes the connection of a rotating perfect fluid with cylindrical symmetry across surfaces $p = 0$ with vacuum and that, moreover, guarantees all positivity conditions in the interior.

Whether Killing vectors of the exterior space of the metrics (3.3) and (4.3) possess rotation or not depends on the algebraic type of $K^m{}_n$ at the surface of the perfect fluid. Despite the rotating perfect fluid the exterior is static in the case A1.

Metrics of the different algebraic types A1 and A2 of $K^m{}_n$ can be transformed by complex coordinate transformations into each other, so the properties of metrics with a different $\operatorname{rank}(K_{mn})$ differ much more than the properties of metrics with different algebraic types of $K^m{}_n$.

For all obtained metrics with the Petrov type D the four-velocity u_α lies in the plane spanned by the two principal null vectors of the Weyl tensor l^α, k^β, i.e., $u^{[\alpha}k^{\beta}l^{\gamma]} = 0$ and the magnetic part of the Weyl tensor with respect to u^α vanishes. As remarked in Carminati and Wainwright[27] the first of these two properties applies to almost all known type D metrics with perfect fluid except those of Wahlquist[28] and Kramer.[29] According to a theorem of Carminati and Wainwright, all metrics with perfect fluid, an equation of state $p = p(\mu)$, the Petrov type D, a vanishing magnetic part of the Weyl tensor with respect to u^α, and $u^{[\alpha}k^{\beta}l^{\gamma]} = 0$ possess one of the following properties: the equation of state obeys $dp/d\mu = 1$ or 0, or at least three Killing vectors exist. The LRS II-metric (3.4)–(3.6) with Petrov type D has four Killing vectors and the other type D metrics satisfy $\mu = p$. From these line elements the Gödel solution is an example of a metric with a one-parameter manifold of families of flat slices. From the type D metrics, (3.18) is, from the mathematical point of view and as a cosmological solution, the most interesting one. As far as the author knows and according to an analysis of Wainwright,[23,24] so far no spatial inhomogeneous metric with rotating and expanding perfect fluid is known. Besides these properties, the metric (3.18) has only one Killing vector. Inhomogeneous metrics with as-general-as-possible kinematic properties are necessary for the theoretical discussion of many aspects of the early universe so the derivation of such metrics became a more-pursued task during the last years.

For performing the computations especially for determining the Killing vectors, computer programs for decoupling and integrating systems of partial differential equations[34] were used.

Concluding, it shall be emphasized that the approach to assume special timelike hypersurfaces of constant exterior curvature K_{ab} and to align coordinates not to Killing vectors but to the structure of $K^a{}_b$, which proved to be fruitful, is not only restricted to metrics with flat hypersurfaces.

ACKNOWLEDGMENTS

I am grateful to H. Stephani, D. Kramer, and G. Neugebauer for helpful discussions, especially to H. Stephani for steady encouragement. Further I want to thank the Jenaer Theoreticians for the warm and stimulating atmosphere.

APPENDIX A: CHRISTOFFEL SYMBOLS FOR GENERAL g_{ab}, N_a, AND N

With the tensor of exterior curvature $K_{\alpha\beta}$ of hypersurfaces $x^4 = \text{const}$ in an arbitrary space-time with the metric

$$ds^2 = g_{ab}(dx^a + N^a dx^4)(dx^b + N^b dx^4) + \epsilon N^2 (dx^4)^2,$$

the Christoffel symbols read

$$\Gamma^4_{bc} = \epsilon(N_{b\|c} + N_{c\|b} - g_{bc,4})/(2N^2) = \epsilon K_{bc}/N,$$

$$\Gamma^a_{bc} = \overset{3}{\Gamma}{}^a_{bc} - \epsilon N^a K_{ac}/N,$$

$$\Gamma^a_{b4} = \overset{3}{\Gamma}{}^{ad}(g_{db,4} + N_{d,b} - N_{b,d})/2$$

$$- N^a N_{,b}/N - \epsilon N^a K_{b4}/N,$$

$$\Gamma^4_{b4} = N_{,b}/N + \epsilon K_{b4}/N,$$

$$\Gamma^a_{44} = \overset{3}{g}{}^{ab}(N_{b,4} - N^n N_{n,b} - \epsilon N N_{,b} + N^n N^d g_{nd,b}/2)$$

$$- N^a(N_{,4} + N^b N_{,b})/N - \epsilon N^a K_{44}/N,$$

$$\Gamma^4_{44} = N_{,4}/N + N^b N_{,b}/N + \epsilon K_{44}/N,$$

with $K_{a4} = K_{ab}N^b$, $K_{44} = K_{ab}N^a N^b$.

APPENDIX B: CONSTRUCTION OF A METRIC WITH $\mu \geq p \geq 0$ IN THE INTERIOR AND WITH CONNECTION TO VACUUM

In this Appendix we will show that in the metric (4.3) are contained specializations describing space-time with perfect fluid, a free equation of state, and generally $\mu \geq p \geq 0$ connected with a vacuum across surfaces $p = 0$.

For a vacuum it follows from (4.4a)–(4.4k) with an appropriate x^4-translation,

$$K^a{}_a = -1/x^4, \quad b = b_0/x^4, \quad b_0 = \text{const}, \quad 0 < 3b_0^2 - 2b_0.$$

To show that functions $K^a{}_a(x^4)$ and $b(x^4)$ exist that obey the inequalities (4.4a)–(4.4c), including the boundary conditions

$$K^a{}_a(x_0^4) = -1/x_0^4, \quad b(x_0^4) = b_0/x_0^4, \quad 0 < 3b_0^2 - 2b_0,$$

we introduce functions $K(x^4), B(x^4)$ with

$$K^a{}_a = K - 1/x^4, \tag{B1a}$$

$$b = B(K - 1/x^4), \tag{B1b}$$

$$K(x_0^4) = 0, \tag{B1c}$$

$$B'(x_0^4) = 0. \tag{B1d}$$

The inequalities (4.4a)–(4.4c) then read

$$0 \le (\mu - p)/2 = K' - K^2 + 2K/x^4 + 2p, \tag{B2}$$

$$0 < p/(K - 1/x^4)^2 + 3B^2 - 2B, \tag{B3}$$

$$0 \le B'(1/x^4 - K) - B(K' - K^2 + 2K/x^4) - p. \tag{B4}$$

The method will be to determine K' step by step for one value of x^4 depending on K at this value of x^4 and also B' and then to repeat this for lower x^4. Because the conditions for K and B,

$$K < 0, \quad K' > 0 \quad \text{for} \quad x^4 < x_0^4$$
$$\text{and} \quad K = 0 \quad \text{for} \quad x^4 = x_0^4$$
$$(\text{resp. } B < 0, \quad B' > 0 \quad \text{for} \quad x^4 < x_0^4$$
$$\text{and} \quad B < 0 \quad \text{for} \quad x^4 = x_0^4)$$

are consistent, it is only necessary to assume K' as high as necessary to guarantee for given K at x^4 that (B2) is obeyed by

$$K' - K^2 + \frac{2K}{x^4} \begin{cases} > 0, & \text{for } \mu - 5p > 0, & \text{(B5a)} \\ < 0, & \text{for } \mu - 5p < 0, & \text{(B5b)} \end{cases}$$

according to the desired equation of state. With K and K' (B2) provides $\mu - 5p$ and together with a chosen equation of state μ and p satisfy $\mu \ge p \ge 0$ because the sign of $\mu - 5p$ in (B5) was determined in correspondence with $\mu \ge p \ge 0$ and $\mu = \mu(p)$. The inequality (B3) is satisfied by $B < 0$. With $K < 0$, the condition (B4) can be obeyed by a sufficiently high value of B'. From (B1) with $K, B, K^a{}_a(x^4)$ and $b(x^4)$ result. The remaining steps (4.4d)–(4.4k) then yield the metric.

[1]C. B. Collins and D. A. Szafron, J. Math. Phys. **20**, 2347 (1979).
[2]H. Stephani, Gen. Relativ. Gravit. **14**, 703 (1982).
[3]H. Stephani and T. Wolf, "Perfect fluid and vacuum solutions of Einstein's field equations with flat 3-dimensional slices," in *Axisymmetric Systems, Galaxies and Relativity: Essays Presented to W. B. Bonnor on His 65th Birthday* (Cambridge U.P., Cambridge, 1985).
[4]T. Wolf, J. Math. Phys. **27**, 2354 (1986).
[5]W. Kundt, Proc. R. Soc. London Ser. A **270**, 328 (1962).
[6]W. Kundt and M. Trümper, Akad. Wiss. Lit. Mainz, Abh. Math. Nat. Kl. **12** (1962).
[7]A. Z. Petrov, in *Recent Developments in General Relativity* (Pergamon–FWN, New York, 1962), p. 379.
[8]Ch. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[9]H. Stephani, *General Relativity* (Cambridge U.P., Cambridge, 1985).
[10]H. Goenner and J. Stachel, J. Math. Phys. **11**, 3358 (1970).
[11]G. S. Hall, J. Phys. A **9**, 541 (1976).
[12]D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980).
[13]H. Stephani, Commun. Math. Phys. **4**, 137 (1967).
[14]A. Das, J. Math. Phys. **14**, 1099 (1973).
[15]T. Wolf, to be published in Class. Quantum Gravit.
[16]G. F. R. Ellis and A. R. King, J. Math. Phys. **8**, 1171 (1967).
[17]J. M. Stewart and G. F. R. Ellis, J. Math. Phys. **9**, 1072 (1968).
[18]R. R. Tabensky and A. H. Taub, Commun. Math. Phys. **29**, 61 (1973).
[19]J. Wainwright, J. Phys. A **12**, 2015 (1979).
[20]I. Ozsváth, J. Math. Phys. **6**, (1965).
[21]D. L. Farnsworth and R. P. Kerr, J. Math. Phys. **7**, 1625 (1966).
[22]K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).
[23]J. Wainwright, J. Phys. A **14**, 1131 (1982).
[24]J. Wainwright, "Bibliography: Perfect Fluid Solutions 1980–1984," preprint.
[25]M. D. Pollock and N. Caderni, Mon. Not. R. Astron. Soc. **190**, 509 (1980).
[26]J. D. Barrow, Philos. Trans. R. Soc. London Ser. A **296**, 273 (1980).
[27]J. Carminati and J. Wainwright, Gen. Relativ. Gravit. **17**, 853 (1985).
[28]H. D. Wahlquist, Phys. Rev. **172**, 1291 (1968).
[29]D. Kramer, Class. Quantum Gravit. **1**, L3 (1984).
[30]F. R. Gantmacher, *Matrizenrechung* I (VEB Deutscher Verlag der Wissenschaften, Berlin, 1958).
[31]T. Lewis, Proc. R. Soc. London Ser. A **136**, 176 (1932).
[32]A. Krasiński, Rep. Math. Phys. **14**, 225 (1978).
[33]E. Kasner, Am. J. Math. **43**, 217 (1921).
[34]T. Wolf, J. Comput. Phys. **60**, 437 (1985).

About vacuum solutions of Einstein's field equations with flat three-dimensional hypersurfaces

Thomas Wolf
Friedrich-Schiller-Universität Jena, Sektion Physik, Max-Wien-Platz 1, DDR 6900 Jena, German Democratic Republic

(Received 19 March 1986; accepted for publication 30 April 1986)

The class of vacuum space-times with a family of flat three-slices and a traceless tensor of exterior curvature K_{ab} is examined. Metrics without symmetry and solutions describing gravitational radiation are obtained. It turns out that there is a correlation between rank (K_{ab}) and the Petrov type. Although the resulting solutions are already known, the richness of the class of space-times with flat slices becomes obvious. An example is given of a metric with one-parameter manifold of families of flat slices.

I. INTRODUCTION

In a series of papers, Collins and Szafron[1] proposed to impose restrictions on submanifolds of space-time to find new exact solutions of Einstein's field equations. The four-velocity u^α of perfect fluid, e.g., should be orthogonal to hypersurfaces with inner symmetries.

In the papers of Stephani[2] and Stephani and Wolf[3] instead of restricting the hypersurfaces to be spacelike or to demand that u^α has no rotation, it was assumed that the inner curvature of the slices $x^4=\text{const}$ vanishes and the tensor of exterior curvature vanishes, too, or is proportional to the three-metric. It is promising to investigate further classes with flat slices because new solutions were obtained as well as the property that metrics with spherical symmetry and satisfying a certain positivity condition possess two families of flat slices. For static perfect fluid this inequality guarantees that the mass function $m(r)$, defined by

$$m(r)=\frac{r}{2}\,(1-e^{-2\lambda(r)})=\frac{1}{2}\int_0^r \mu(x)x^2\,dx\,,$$

is positive (with $e^{2\lambda}=g_{rr}$ in the diagonal form of the metric). An example of a metric with three Killing vectors acting on two-surfaces, which not only has two families but a one-parameter manifold of families of flat slices, is given by

$$ds^2=(dx^1)^2+(dx^2)^2+e^{2\lambda}(dx^3)^2-e^{2\nu}(dx^4)^2,$$
$$\lambda(x^3,x^4),\nu(x^3,x^4).$$

Every parametrization by

$$\bar{x}^4=\bar{x}^4(x^3,x^4),\quad \bar{x}^4{}_{,\alpha}\bar{x}^{4,\alpha}\neq 0\,, \tag{1.1}$$

provides a family of flat slices. Since $\bar{\nu}_{,3}=0$ can be attained and by this we can obtain a geodesic normal vector as well as $\bar{\nu}_{,3}\neq 0$ and with it a nongeodesic normal vector, the slicings are invariantly distinct and cannot be transformed into each other by symmetry transformations. Because a new parametrization $\bar{x}^4=\bar{x}^4(x^4)$ gives the same family, it follows from (1.1) that in every point of space-time there exists a one-parameter manifold of normal vectors to flat slices.

II. THE METRIC AND THE FORMALISM

A family of hypersurfaces is given that is parametrized by x^4. Because these hypersurfaces may be spacelike or timelike, x^4 is a timelike or spacelike coordinate. With the normal vector $n_\alpha=(0,0,0,N)$, $n^\alpha n_\alpha=\epsilon=\pm 1$, the three-metric $g_{ab}(x^\alpha)$ within the hypersurfaces, and the three-vectors N^a, which fix the mutual position of coordinate systems of neighboring slices, the full space-time metric reads

$$ds^2=g_{ab}(dx^a+N^a\,dx^4)(dx^b+N^b\,dx^4)+\epsilon(N\,dx^4)^2\,. \tag{2.1}$$

In the following, Latin indices are moved by g_{ab} and $\overset{3}{g}{}^{ab}$, i.e.,

$$N_a=g_{ab}N^b,\quad g_{ab}\overset{3}{g}{}^{bc}=\delta^c_a\,.$$

Here $g^{\alpha\beta}$ and n^α take the form

$$g^{\alpha\beta}=\begin{pmatrix}\overset{3}{g}{}^{ab}+\epsilon N^aN^b/N^2 & -\epsilon N^b/N^2\\ -\epsilon N^a/N^2 & \epsilon/N^2\end{pmatrix},$$
$$n^\alpha=(-N^a/N,1/N)\,.$$

The tensor of exterior curvature $K_{\alpha\beta}$ is defined to be the negative projection of the gradient of n_α onto the hypersurface

$$K_{\alpha\beta}=n_{\alpha;\gamma}(\delta^\gamma{}_\beta-\epsilon n^\gamma n_\beta)\,.$$

The symmetry $K_{\alpha\beta}=K_{\beta\alpha}$ results from the vanishing of the rotation of n_α. Because n^α is normalized, $K_{\alpha\beta}$ satisfies $K_{\alpha\beta}n^\beta=0$, i.e., $K_{4a}=N^bK_{ba}$, $K_{44}=N^aN^bK_{ab}$. With the covariant three-derivative $_{\|a}$ we eventually get

$$K_{ab}=(N_{a\|b}+N_{b\|a}-g_{ab,4})/2N\,. \tag{2.2}$$

Applying projection techniques described in Ref. 4, § 21 and Ref. 5, the space-time curvature tensor can be expressed by K_{ab}, n_α, their covariant derivatives and $\underset{n}{\pounds}K_{ab}$. We prefer the following form for the Ricci tensor:

$$\epsilon NR_{ab}=g_{r(b}K^r{}_{a),4}-K_{ab\|r}N^r-K_{ab}K^r{}_rN$$
$$+K^r{}_bN_{[a,r]}+K^r{}_aN_{[b,r]}-N_{,b\|a}\,, \tag{2.3a}$$

$$\epsilon NR^4{}_m=K^b{}_{b\|m}-K^b{}_{m\|b}\,, \tag{2.3b}$$

$$\epsilon(R-2R_{ab}\overset{3}{g}{}^{ab})=K^a{}_a{}^2-K^a{}_bK^b{}_a\,. \tag{2.3c}$$

The main advantage of the restriction $K^a{}_a=0$ consists of the linearization of (2.3a) with respect to K_{ab}.

The approach will be to determine the decomposition of K_{ab} in eigenvectors by analyzing (2.3c) and $K^a{}_a = 0$. Equation (2.3b) will provide the vanishing of rotation of a null three-vector invariantly defined by K_{ab}. In the next step a flat three-metric g_{ab} will be obtained, which is aligned to this null vector and simplified as much as possible by the remaining coordinate transformations.

We do not replace the components K_{ab} in the field equations (2.3a) and (2.3b) by the expression (2.2) but treat them as independent metric functions and expose them to further restrictions. By doing this Eq. (2.2) no longer have the status of definition equations but of a tensor equation to be solved in addition to the field equations. Therefore they are called K_{ab}-equations in the following. In the three-metric adapted to the problem these equations together with the remaining field equations (2.3a) and (2.3b) will be solved.

III. A SUITABLE NULL TRIAD

With $0 = K^a{}_b K^b{}_a$ from (2.3c) the flat slices must be timelike because otherwise the case $0 = K_{ab}$ described by Stephani[1] would follow. Expanded in a null triad k_a, l_a, q_a with

$$-k_a l^a = q_a q^a = 1, \quad k_a k^a = l_a l^a = q_a k^a = q_a l^a = 0,$$

K_{ab} reads

$$K_{ab} = Ak_a k_b + Bl_a l_b + Cq_a q_b + 2Dk_{(a}q_{b)} + 2Ek_{(a}l_{b)} + 2Fl_{(a}q_{b)} .$$

The algebraic conditions $0 = K^a{}_a = K^a{}_b K^b{}_a$ give

$$0 = C - 2E \tag{3.1a}$$

and

$$0 = AB + 3E^2 - 2DF. \tag{3.1b}$$

With real parameters a,b,c the yet-possible transformations of the null triad are

$$k_m = ak_m', \quad l_m = (1/a)l_m', \quad q_m = q_m', \tag{3.2a}$$

$$\begin{aligned} &l_m = l_m', \quad q_m = q_m' + bl_m', \\ &k_m = k_m' + bq_m' + (b^2/2)1_m', \end{aligned} \tag{3.2b}$$

$$\begin{aligned} &k_m = k_m', \quad q_m = q_m' + ck_m', \\ &l_m = l_m' + cq_m' + (c^2/2)k_m' \end{aligned} \tag{3.2c}$$

and yield, e.g., for $a = c \neq b$, the new components

$$\begin{aligned} &A' = A, B' = b^4 A/4 + b^3 D + b^2(C+E) + 2bF + B, \\ &C' = b^2 A + 2bD + C, \\ &D' = bA + D, \quad E' = b^2 A/2 + bD + E, \\ &F' = b^3 A/2 + 3b^2 D/2 + b(C+E) + F. \end{aligned} \tag{3.3}$$

These transformations are to be applied to transform some of the components to zero. It turns out that for $A \neq 0 \neq B$ the conditions (3.1) guarantee the existence of a real parameter b to achieve $B' = 0$. For $A \cdot B = 0$ as well as $A = B = 0$ transformations can easily be found to obtain further $C = E = DF = 0$. The result is that K_{ab} can be written as

$$K_{ab} = Ak_a k_b + 2Fl_{(a}q_{b)}, \tag{3.4}$$

renaming $k_a \leftrightarrow l_a$ if necessary.

From the field equations $0 = K^b{}_{m\|b}$, it follows that with $F = 0$, $K^b{}_{m\|b} q^m = 0$ and therefore $k_{[a\|b}k_{c]} = 0$, and with $A = 0$, $K^b{}_{m\|b} l^m = 0$ and therefore $l_{[a\|b}l_{c]} = 0$. Because rotation vanishes, the vectors k_a (resp. l_a) are proportional to a gradient and by a transformation $k_m' = ak_m$, $l_m' = (1/a)l_m$ are equal to the gradient of a function u. In the case $A \cdot F \neq 0$, it cannot be derived for any vector of (3.4) that its rotation vanishes. Because consequently no appropriate three-metric was available, this case could not be solved in general, so this paper is constrained to $A \cdot F = 0$, i.e., rank $(K_{mn}) < 3$. In the summary a special solution for $A \cdot F \neq 0$ is given.

Having the algebraic structure of K_{ab} established suitable flat three-metrics are to be found as follows.

IV. APPROPRIATE THREE-METRICS

We want to introduce a coordinate system that is at most adapted to the preferred null vector. To this aim we choose $u = x^3$ as a coordinate and demand $k_a = (0,0,1)$. It can be shown easily that coordinates $(x = x^1, v = x^2)$ are attainable that preserve $K^a = (0,1,0)$ and that give the three-metric the form

$$ds^2 = g_{11}\,dx^2 + 2g_{13}\,dx\,du + 2\,dv\,du + g_{33}\,du^2 . \tag{4.1}$$

The remaining coordinate transformations are $\bar{u} = \bar{u}(u)$, $\bar{x} = \bar{x}(x,u)$, and $\bar{v} = v/\bar{u}_{,u} + f(x,u)$, including the—for the time uninteresting—x^4-dependences. Analyzing the flatness property for (4.1) (see Appendix A), which has not been taken into consideration up to now, three cases occur corresponding to different structures of $k_{a\|b}$:

$$(1)\ K_{a\|b} \sim k_a k_b \leftrightarrow g_{11} = 1, \quad g_{13} = g_{33} = 0; \tag{4.2a}$$

$$\begin{aligned} &(2)\ k_{[m}k_{a]\|b} = 0 \text{ or } k_{a\|[b}k_{m]} \neq 0, k^a{}_{\|a} = 0 \\ &\qquad \leftrightarrow g_{11} = 1, \quad g_{13} = -2v/x, \\ &\qquad g_{33} = p(u)x^2 + q(u)x; \end{aligned} \tag{4.2b}$$

$$\begin{aligned} &(3)\ k^a{}_{\|a} \neq 0 \\ &\qquad \leftrightarrow g_{11} = v^2, \quad g_{13} = l(x,u)v^2 + m(x,u) \\ &\qquad g_{33} = l^2 v^2 - 21_{,x} v + 2lm + 2n(x,u) . \end{aligned} \tag{4.2c}$$

Instead of illustrating all six cases of the metrics (4.2) with $A = 0$ (resp. $F = 0$) only the case (4.2a) with $A = 0$ shall be explained in Appendix B. The results for all six cases are given in the summary.

V. SUMMARY

The subjects of this paper are the vacuum solutions of Einstein's field equations with a family of flat slices and a corresponding tensor of exterior curvature K_{ab} that is traceless and has a rank lower than 3. Utilizing the field equation (2.3c), it could be shown that with vectors k_a, l_a, and q_a of an appropriate null triad, K_{ab} takes the form

$$K_{ab} = Ak_a k_b + 2Fl_{(a}q_{b)} .$$

Because the field equations $0 = R^4{}_a$ yield for $F = 0$ (resp. $A = 0$) the vanishing of the rotation of k_a (resp. l_a), all further field equations could be solved by introducing appropriate coordinates and adapted three-metrics. These metrics correspond to the cases

(1) $k_{a\|b} \sim k_a k_b$;

(2) $0 \neq k_{[m} k_{a]\|b}$ or $0 \neq k_{a\|[b} k_{m]}$, $k^a{}_{\|a} = 0$;

(3) $k^a{}_{\|a} \neq 0$ with $F = 0$ and as well in analogy to the three cases with $A = 0$ and the same subdivision for l_a instead of k_a. All resulting solutions are subcases of already known metrics:

I. $K_{ab} = A k_a k_b$,

I.1. $k_{a\|b} \sim k_a k_b$.

Under these assumptions two metrics exist, both belonging to the *pp*-waves, because $\xi^\alpha = \delta^\alpha_2$ is a covariant constant null Killing vector:

$$ds^2 = dx^2 + 2\, dv\, du + [2G_{,4}\, dx + 2(w - xG_{,4u})du + (G_{,4}{}^2 + G_{,4uu}{}^2/F^2)dx^4]dx^4 .$$

Here, $G(u,x^4)$ and $F(u)$ are arbitrary functions, whereas $w(u,x^4)$ has to be calculated from

$$w_{,u} = -G_{,4uu}[G - (1/F)(G_{,uu}/F)_{,uu}] .$$

Now,

$$ds^2 = dx^2 + 2\, dv\, du + [2G_{,44}\, dx + 2(w - xG_{,44u})du + (G_{,44}{}^2 + (x + G_{,4})^2)dx^4]dx^4 .$$

The $G(u,x^4)$ and $w(u,x^4)$ have to be calculated from

$$0 = G_{,uu44} + G_{,uu}, \quad 0 = F + G_{,4}G_{,uu} - w_{,u} ,$$

with the arbitrary function $F(u)$.

Another solution is

I.2. $k_{[m} k_{a]\|b} \neq 0$ or $k_{a\|[b} k_{m]} \neq 0$, $k^a{}_{\|a} = 0$.

For the following solution $k^\alpha = \delta^\alpha_2$ is a nontwisting nondiverging principal null direction of the Weyl tensor of multiplicity 4. This solution therefore belongs to the Petrov type N metrics of Kundt's class[6,7]:

$$\begin{aligned} ds^2 = dx^2 &+ 2\, dv\, du - 4v/dx\, du \\ &+ (v^2/x^2 + px^2 + qx)du^2 \\ &+ [-4vj/x\, dx + 2jdv + 2(v^2j/x^2 - vj_{,m} \\ &+ (j_{,uu} + jp)x^2 + q\, jx)du \\ &+ (1 - jj_{,u}v + jj_{,uu}x^2)dx^4]dx^4 , \end{aligned}$$

where $V(u,x^4)$ and $F(V)$ are arbitrary functions, and $j(u,x^4)$, $p(u,x^4)$, and $q(u,x^4)$ have to be calculated from

$$j = V_{,4}/V_{,u} ,$$

$$(-2j_{,u} + \partial_4 - j\,\partial_u)q = V_{,u}{}^2 F ,$$

$$(-2j_{,u} + \partial_u - j\,\partial_u)p = 2j_{,uuu} + V_{,u}{}^2\overline{F}, \quad \overline{F} = 1 \text{ or } 0 .$$

For

I.3. $k^a{}_{\|a} \neq 0$,

no vacuum solutions exist.

Other solutions are

II. $K_{ab} = 2Fl_{(a}q_{b)}$,

II.1. $l_{a\|b} \sim l_a l_b$ (the case in Appendix B.).

In addition to not explicitly derived subcases of the *pp*-waves the following solution is contained for which $k^\alpha = \delta^\alpha_2$ is a nontwisting, nondiverging principal null direction of the Weyl tensor of multiplicity 3. It therefore belongs to the Petrov type III metrics of Kundt's class:

$$\begin{aligned} ds^2 = dx^2 &+ 2\, dv\, du + [2g\, dx + 2j\, dv \\ &+ 2(-vj_{,u} + x^2 j_{,uu} + x(2sJ - g_{,u}) + f)du \\ &+ (g^2 + 2j(-vj_{,u} + x^2 j_{,uu} \\ &+ x(2sJ - g_{,u}) + f) + s^2)dx^4]dx^4 . \end{aligned} \tag{5.1}$$

Whereas for $j_{,uu} = 0$ the metric describes *pp*-waves, for $j_{,uu} \neq 0$ we have to calculate

$$j(u,x^4) \text{ from } 0 = (jj_{,uu}{}^{-1/3})_{,u} - (j_{,uu}{}^{-1/3})_{,4} , \tag{5.2}$$

a special solution V from $0 = V_{,4} - jV_{,4}$, $s(u,x^4)$ from $s = j_{,uu}{}^{4/3}G(V)$ with an arbitrary function $G(V)$, $g(u,x^4)$ from

$$0 = g_{,uu4} - jg_{,uuu} + 2j_{,u}g_{,uu} + j_{,uu}g_{,u} ,$$

$J(u,x^4)$ from

$$0 = (jJ)_{,u} - J_{,4} + j_{,uu}g/s ,$$

$f(u,x^4)$ from

$$0 = f_{,u4} - jf_{,uu} + 2j_{,u}f_{,u} + j_{,uu}f + 2sJ(sJ - g_{,u}) - g(2sJ - g_{,u})_{,u} - ss_{,uu} .$$

Except the equation for calculation of j all equations are linear. Two special solutions of (5.2) are given by

$$\text{(a) } j_{,4} = 0, \quad u = \int (c_1 j^4 + c_2)^{-1/2}\, dj ,$$

$$\text{(b) } \hat{j} = \hat{j}(u), \quad c_1 u = \int [\hat{j}(\hat{j}^{c_2} + c_3\hat{j}^{-c_2})^2]^{-1/4}\, d\hat{j} , \tag{5.3}$$

$$j = \frac{\hat{j}^{3/4}}{3x^4 c_1[\hat{j}^{c_2} + c_3\hat{j}^{-c_2}]^{1/4}} ,$$

with c_i = const and the choice $c_1 = 1$ corresponds to a coordinate transformation $u = u{\cdot}c$, $v = v/c$. It shall be remarked that this metric has, in general, no Killing vector, as a straightforward calculation of case (5.3) shows.

For solutions

II.2., $l_{[m} l_{a]\|b} = 0$ or $l_{a\|[b} l_{m]} \neq 0$, $l^a{}_{\|a} = 0$,

II.3., $l^a{}_{\|a} \neq 0$,

no vacuum solutions exist.

Under the assumptions $K^a{}_a = 0$, rank $(K_{ab}) < 3$, the given distinction of cases is complete. The case K_{ab} regular, i.e., $AF \neq 0$ could not be solved in general. A special solution for nontwisting k_a with $k_{a\|b} \sim k_a k_b$ is

$$ds^2 = dx^2 + 2\, dv\, du + [2v\, dx + 2x\, dv + 2cu\, du + (1 + v^2 + 2cxu)dx^4]dx^4 .$$

This solution has the Petrov type I and four Killing vectors. It is equivalent to a metric of Petrov[8]

$$K^2\, ds^2 = dx^2 + e^{-2x}dy^2 + e^x(\cos(\sqrt{3}x)(dz^2 - dt^2) - 2\sin(\sqrt{3}x)dz\, dt) ,$$

but is obviously formulated in more suitable coordinates.

As could have been expected a dependence of the Petrov type on the rank (K_{ab}) was established because for $K^a{}_a = 0$ and $K_{ab} = Ak_a k_b$ with rank 1 only type N solutions exist, for

$K_{ab}=2Fl_{(a}q_{b)}$ with rank 2, type III solutions occur, and for K_{ab} of rank 3, the Petrov solution with Petrov type I is contained.

A conjecture of Berger *et al.*[9] and Wainwright[10] not formulated in every detail says that there is a correlation between the nature of existing gravitational waves and the Cotton–York tensor, which describes the conformally invariant part of the curvature of the hypersurfaces. This hypothesis arose from the fact that the Szekeres solution contains conformally flat slices with vanishing Cotton–York tensor and that, on the other hand, these metrics do not possess gravitational radiation, as was shown by Bonnor.[11] Because we derived type N metrics (e.g., *pp*-waves) with flat timelike slices, in the conjecture at least the nature of the hypersurfaces has to be fixed more precisely.

As a further interesting result it can be established that for vacuum solutions the existence of flat slices is not coupled to the existence of symmetries as the metric (5.1) shows. Moreover flat slices excluded neither algebraically special nor algebraically general metrics. Even if no new metrics were derived the independence of the approach based on flat slices is underlined by the richness of the obtained solutions.

In contrast to the pure algebraic restrictions on K_{ab}, in addition to the demand of flatness in this article, in the preceding paper,[12] we make restrictions of a differential kind on K_{ab}. A result of this will be that the obtained metrics possibly have the Petrov type I and on the other hand, do possess at least one Killing vector.

ACKNOWLEDGMENTS

I would like to thank H. Stephani and D. Kramer for many comments and suggestions.

APPENDIX A: ANALYSIS OF $\overset{3}{R}{}^a{}_{bcd}=0$ FOR g_{ab}

Starting with the form of the three-metric

$$\overset{3}{ds^2}=g_{11}\,dx^2+2g_{13}\,dx\,du+2\,dv\,du+g_{33}\,du^2,$$

$$x^1=x,\quad x^2=v,\quad x^3=u\,,$$

We obtain g_{11} from $0=\overset{3}{R}{}^3{}_{121}$ and a transformation $\bar v=v+f_1(x,u)$ [resp. $\bar x=\bar x(x,u)$] yields $g_{11}=v^2$ or $g_{11}=1$. In the case $g_{11}=1$ we obtain g_{13} from $0=\overset{3}{R}{}^3{}_{113}$, while in the other case we obtain it from $0=\overset{3}{R}{}^3{}_{123}$. With this, g_{33} can be determined from the residual equations. The result is that a metric of a three-space with one time dimension allowing a null vector k^a with $k_{[a\|b]}$ to have the form $k^a=\delta^a_2\,k^2$, $k_a=\delta^3_a k_3$, i.e., $k_3=k_3(u)$, is transformable in one of the following three forms. The characterization of the three cases by $k_{a\|b}$ is invariant against transformations of the null vectors $k^{n'}=ak^n$, $l^{n'}=1/al^n$, with $a=a(u)$ because of $k_{[b\|c]}=0$ and $k_b=\delta^3{}_b k_3$:

$$\overset{3}{ds^2}=g_{11}\,dx^2+2g_{13}\,dx\,du+2\,dv\,du+g_{33}\,du^2\,,$$

$$x^1=x,\quad x^2=v,\quad x^3=u\,.$$

The three cases are

(1) $k_{a\|b}\sim k_a k_b\leftrightarrow g_{11}=1,\quad g_{13}=g_{33}=0,$

$$\Gamma^a_{bc}=0,\quad \overset{3}{g}{}^{ab}=\begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix};$$

(2) $k_{[m}k_{a]\|b}\neq 0$ or $k_{a\|[b}k_{m]}\neq 0,\quad k^a{}_{\|a}=0$

$$\leftrightarrow g_{11}=1,\quad g_{13}=-2v/x,\ g_{33}=p(u)x,$$

$$\overset{3}{g}{}^{ab}=\begin{pmatrix}1&2v/x&0\\2v/x&3v^2/x^2-px^2-qx&1\\0&1&0\end{pmatrix},$$

$$\Gamma^1_{13}=\Gamma^2_{11}=\frac{2v}{x^2},\quad \Gamma^1_{23}=\Gamma^2_{12}=-\Gamma^3_{13}=\frac{-1}{x},$$

$$\Gamma^1_{33}=-\left(px+\frac{q}{2}+\frac{v^2}{x^3}\right),\quad \Gamma^2_{13}=\frac{2v^2}{x^3}-\frac{q}{2},$$

$$\Gamma^2_{32}=\Gamma^3_{33}=-v/x^2,$$

$$\Gamma^2_{33}=\frac{p_{,u}x^2}{2}+\frac{q_{,u}x}{2}-\frac{v^3}{x^4}-pv;$$

(3) $k^a{}_{\|a}\neq 0\leftrightarrow g_{11}=v^2,\quad g_{13}=l(x,u)v^2+m(x,u),$

$$g_{33}=l^2v^2-2l_{,x}v+lm+2n(x,u)\,,$$

$$\overset{3}{g}{}^{ab}=\begin{pmatrix}1/v^2&-l-m/v^2&0\\-l-m/v^2&m^2/v^2+2l_{,x}v-2n&1\\0&1&0\end{pmatrix},$$

where l,m,n still have to satisfy the conditions

$$0=-m_{,u}+2l_{,x}m+lm_{,x}+n_{,x}\,,$$

$$0=-n_{,u}+2l_{,x}n+ln_{,x}-l_{,xxx}\,,$$

$$\Gamma^1_{11}=lv+\frac{m}{v},\quad \Gamma^1_{12}=\frac{1}{v},\Gamma^1_{13}=l^2v+\frac{lm}{v},$$

$$\Gamma^1_{23}=\frac{1}{v},\quad \Gamma^1_{33}=l_{,u}-2ll_{,x}+\frac{l_{,xx}}{v}+l^3v+\frac{l^2m}{v},$$

$$\Gamma^2_{11}=m_{,x}+2nv-m^2/v-l_{,x}v^2,$$

$$\Gamma^2_{23}=-l_{,x}-lm/v,$$

$$\Gamma^2_{33}=-l^2l_{,x}v^2+v(-ll_{,xx}+2l^2n+2l_{,x}{}^2-l_{,xu})-2l_{,x}n+lm_{,u}+n_{,u}-ml_{,xx}/v-l^2m^2/v,$$

$$\Gamma^3_{11}=-v,\quad \Gamma^3_{13}=-lv,\quad \Gamma^3_{33}=l_{,x}-l^2v.$$

The given Γ^a_{bc} are three-Christoffel symbols.

APPENDIX B: THE CASE $K_{ab}=2Fl_{(a}q_{b)}$, $l_{a\|b}\sim l_a l_b$

The subject of this appendix are the field equations

$$0=g_{r(b}K^r{}_{n),4}-K_{nb\|a}N^a+K^r{}_bN_{[n,r]}+K^r{}_nN_{[b,r]}-\epsilon N_{,b\|n}\,,\tag{B1}$$

$$0 = K^{b}{}_{n\|b}\ , \tag{B2}$$

together with

$$K_{ab} = (N_{a\|b} + N_{b\|a} - g_{ab,4})/2N \tag{B3}$$

and the algebraic conditions

$$K_{2a} = K_{11} = 0\ , \tag{B4}$$

because of $K_{ab} = 2Fl_{(a}q_{b)}$ and $l_a = \delta^3{}_a l_3$, $l^a = \delta^a_2\, l^2$, $l^a q_a = 0 = q_2$.

In all six cases the approach consists of the calculation of N_a, N from the K_{ab}-equation (B3) starting with the v-integration for $(a,b) = (2,2)$, (1.2) and the introduction of functions of the variables x,u,x^4. It follows the comparison of v-coefficients in $(a,b) = (2,3)$, $(1,1)$ and x-integration of the resulting equations. By that the functions of variables x,u,x^4 can be expressed by functions of the variables u,x^4. Also the field equations (B1) are investigated in the sequence $(n,b) = (2,b)$, $(1,1)$,... . By v-integration and comparison of v-coefficients and x-integration and comparison of x-coefficients only equations for functions of u and x^4 remain to be solved. For the case (4.2a), $A = 0$, to be illustrated here it follows, from (B3), (B4) with $(a,b) \neq (1,3)$, $(3,3)$, and the functions k, g, j, w, that

$$N_1 = vk(u,x^4) + g(u,x^4), \tag{B5a}$$

$$N_2 = xk + j(u,x^4)\ , \tag{B5b}$$

$$N_3 = vxk_{,u} - vj_{,n} + w(x,u,x^4)\ . \tag{B5c}$$

Before tackling the remaining equations [(B3) for $(a,b) = (1,3)$, $(3,3)$], N can be obtained from (B1) and $(n,b) \neq (1,3)$, $(3,3)$ with functions $d,h,s,$

$$N = vd(u,x^4) - x^2 d_{,u} + xh(u,x^4) + s(u,x^4)\ , \tag{B6}$$

with the resulting condition

$$0 = K_{13}k - d_{,u}\ . \tag{B7}$$

The field equations $0 = K^{b}{}_{m\|b}$ are equivalent to

$$\begin{aligned} 0 &= K_{13,v}\ , \\ 0 &= K_{33,v} + K_{13,x}\ , \end{aligned} \tag{B8}$$

and taking with N_a, N from above the form

$$0 = -NK_{13,v} = k_{,u} + K_{13}d\ , \tag{B9}$$

and with the conclusion $0 = K_{13,x}d$,

$$\begin{aligned} 0 &= NK_{33,v} + NK_{13,} \\ &= xk_{,uu} - j_{,uu} - K_{33}d + (xh + s)K_{13,x}\ . \end{aligned} \tag{B10}$$

Eliminating K_{33} with this in the field equation (B1), $(n,b) = (1,3)$ [written down under consideration of $K_{13,x}k = 0$ from (B7)]:

$$\begin{aligned} 0 = {} & [K_{13}(xk - j)]_{,u} + K_{33}k - K_{13,x}g \\ & + 2xd_{,uu} - h_{,u} + K_{13,4}\ , \end{aligned} \tag{B11}$$

and assuming $k \neq 0 \rightarrow K_{13,x} = 0$ from (B7), so the comparison of x-coefficients yields

$$0 = kk_{,uu} + dd_{,uu}\ .$$

On the other hand, (B7) and (B9) provide

$$0 = kk_{,u} + dd_{,u}\ .$$

After differentiation of this equation with respect to u and considering the above condition we get $0 = k_{,u}{}^2 + d_{,u}{}^2$ and with (B7) a contradiction to $k \neq 0$. Therefore it follows that $k = 0$ and, with (B9), $d = 0$.

With

$$N_1 = g(u,x^4)\ , \tag{B12a}$$

$$N_2 = j(u,x^4)\ , \tag{B12b}$$

$$N_3 = -vj_{,u} + w(x,u,x^4)\ , \tag{B12c}$$

$$N = xh(u,x^4) + s(u,x^4)\ , \tag{B12d}$$

$$K_{13} = (g_{,u} + w_{,x})/(2xh + 2s)\ , \tag{B12e}$$

$$K_{33} = (-vj_{,uu} + w_{,u})/(xh + s)\ , \tag{B12f}$$

the equations $0 = R_{13} = R_{33}$, i.e., (B1) with $(n,b) = (1,3)$, $(3,3)$ and (B8), i.e.,

$$0 = -j_{,uu}(hx + s) + K_{13,x} \tag{B13}$$

remain to be solved. Doing this the cases $h \neq 0$ and $h = 0$ have to be distinguished.

The case $h \neq 0$: By integrating (B13) with respect to x a logarithmic x-dependence of K_{13} is obtained that because of $0 = R_{13}$, results in $j_{,uu} = 0$. By an appropriate coordinate transformation $u = u(\bar{u},x^4)$ linear in $\bar{u}$ we get $\bar{N}_2 = \bar{j} = 0$. Also, without solving the complicated system of differential equations resulting from comparing coefficients of powers of x in $R_{33} = 0$, it is obvious that $\xi^a = \delta^a_2$ is a covariantly constant Killing vector and existing solutions belong to the plane-fronted gravitational waves with parallel rays (*pp*-waves).

The case $h = 0$: To calculate $w(x,u,x^4)$ we replace K_{13} from (B12e) in (B13) and perform two x-integrations. With new functions $J(u,x^4)$, $f(u,x^4)$ the starting point for solving $0 = R_{13} = R_{33}$ is

$$N_1 = g(u,x^4)\ , \tag{B14a}$$

$$N_2 = j(u,x^4)\ , \tag{B14b}$$

$$\begin{aligned} N_3 = {} & -vj_{,u} + x^2 j_{,uu} \\ & + 2x(s(u,x^4)\cdot J(u,x^4) - g_{,u}) + f(u,x^4)\ , \end{aligned} \tag{B14c}$$

$$N = s\ , \tag{B14d}$$

$$K_{13} = j_{,uu}x/2 + J\ , \tag{B14e}$$

$$K_{33} = (-vj_{,uu} + x^2 j_{,uuu} + x(2sJ - g_{,u})_{,u} + f_{,u})/s\ . \tag{B14f}$$

From $0 = R_{13}$, it follows that

$$0 = (jj_{,u}/s)_{,u} - (j_{,uu}/s)_{,4}\ , \tag{B15a}$$

$$0 = (jJ)_{,u} - J_{,4} + j_{,uu}g/s\ , \tag{B15b}$$

whereas $0 = R_{33}$ leads to

$$0 = -\frac{2j_{,u}j_{,uuu}}{s} + \frac{3j_{,uu}{}^2}{s} - j\left(\frac{j_{,uuu}}{s}\right)_{,u} + \left(\frac{j_{,uuu}}{s}\right)_{,4}\ .$$

In the following we assume $j_{,uu} \neq 0$ because, on the other hand, *pp*-waves would occur as in the case $h \neq 0$. Replacing $s_{,4}$ with (B15a) and integrating with respect to u we obtain, with the new function $\gamma(x^4)$,

$$0 = 3j_{,u} - jj_{,uuu}/j_{,uu} + j_{,uu4}/j_{,uu} + \gamma\ .$$

A transformation

$$u = \alpha(x^4)\bar{u},\ v = \bar{v}/\alpha \rightarrow j = N_2 = \bar{N}_2\alpha = \bar{j}\alpha,$$

with $-\alpha_{,4}/\alpha = \gamma$ yields $\bar{\gamma} = 0$ and therefore

$$0 = (jj_{,uu}{}^{-1/3})_{,u} - (j_{,uu}{}^{-1/3})_{,4} . \tag{B16}$$

With a solution j of (B16) we get s from (B15a) with a special solution V of

$0 = V_{,4} - jV_{,u}$ and an arbitrary $G(V)$:

$$s = j_{,uu}{}^{4/3} G(V) .$$

From the last equation $0 = R_{33}$, two equations follow in u and x^4. Replacing $J_{,4}$ in one of them by (B15b) and $j_{,uu4}$ by (B16), J vanishes totally out of it. These equations linear in g (resp. f) may be regarded as determining g from j and f from j,J,s,g.

[1]C. B. Collins and D. A. Szafron, J. Math. Phys. **20**, 2347 (1979).
[2]H. Stephani, Gen. Relativ. Gravit. **14**, 703 (1982).
[3]H. Stephani and T. Wolf, in *Axisymmetric Systems, Galaxies and Relativity: Essays Presented to W. B. Bonnor on his 65th Birthday* (Cambridge U. P., Cambridge, 1985).
[4]Ch. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[5]H. Stephani, *General Relativity* (Cambridge U. P., Cambridge, 1985).
[6]W. Kundt and M. Trümper, Akad. Wiss. Lit. Mainz, Abhandl. Math.Nat. Kl. **12** (1962).
[7]W. Kundt, Proc. R. Soc. London Ser. A **270**, 328 (1962).
[8]A. Z. Petrov, in *Recent Developments in General Relativity* (Pergamon–PWN, New York, 1962), p. 379.
[9]B. K. Berger, D. M. Eardley, and D. W. Olson, Phys. Rev. D **16**, 3086 (1977).
[10]J. Wainwright, J. Phys. A: Math. Gen. **12**, 2015 (1979).
[11]W. Bonnor, Commun. Math. Phys. **51**, 191 (1976).
[12]T. Wolf, J. Math. Phys. **27**, 2340 (1986).

An example of affine collineation in the Robertson–Walker metric

M. L. Bedran[a)]
Centro de Estudios Nucleares, Universidad Nacional Autonoma de México, Circuito Exterior, C.U., 04510 México, D. F., Mexico

B. Lesche
Instituto de Física, Universidad Nacional Autonoma de México, Cuernavaca, Mexico

(Received 18 December 1985; accepted for publication 30 April 1986)

An affine collineation for the Robertson–Walker metric is found. It implies a condition on the metric that is compatible with Einstein's equations for a perfect fluid satisfying the Hawking–Ellis energy conditions. It is shown how the geodesics of the metric are obtained from the constant of motion associated to the affine collineation.

I. INTRODUCTION

Affine collineations are symmetries of affine spaces defined by the vanishing Lie derivative of the affine connections[1]

$$\underset{\xi}{£}\Gamma^{\alpha}_{\gamma\beta} = \xi^{\alpha}{}_{;\beta;\gamma} + R^{\alpha}{}_{\beta\gamma\sigma}\xi^{\sigma} = 0, \tag{1.1}$$

where ; denotes covariant derivative and $R^{\alpha}{}_{\beta\gamma\sigma}$ is the Riemann tensor.In Riemannian spaces Eq. (1.1) is equivalent to

$$\xi_{(\alpha;\beta);\gamma} = 0, \tag{1.2}$$

where the ( ) indicate symmetrization in the indices α, β. Special cases of affine collineations in Riemannian spacetimes with metric $g_{\alpha\beta}$ are homothetic motions

$$\xi_{(\alpha;\beta)} = \text{const} \times g_{\alpha\beta}. \tag{1.3}$$

and Killing vectors

$$\xi_{(\alpha;\beta)} = 0. \tag{1.4}$$

Affine collineations are transformations that keep the geodesics of space-time unchanged, although they may change the space-time metric. Hojman *èt al.*[2] showed that affine collineations are non-Noetherian symmetries and constructed new constants of motion associated to them.

An example of affine collineation was given by Katzin and Levine[3] in a two-dimensional affine space, which was not a metric space. In this paper we give an example of this type of symmetry in the Robertson–Walker metric and show that the existence of the affine collineation imposes one condition on the metric that is compatible with Einstein's equations for a perfect fluid distribution of matter. The equation of state thus obtained satisfies all the requirements for a reasonable physical system.

Finally we show how the geodesics of the Robertson–Walker metric can be obtained from the constant of motion associated to the affine collineation.

II. METRIC, AFFINE COLLINEATION, AND FIELD EQUATIONS

The Robertson–Walker metric in spherical coordinates reads

$$ds^2 = dt^2 - R^2(t)\left(\frac{dr^2}{1-kr^2} + r^2\,d\theta^2 + r^2\sin^2\theta\,d\varphi^2\right), \tag{2.1}$$

where $k = 0, \pm 1$. We shall look for a vector ξ_α that satisfies Eq. (1.2). In this paper we consider a vector in the direction of time given by

$$\xi_\alpha = f(t)\delta^0_\alpha, \tag{2.2}$$

which yields

$$\xi_{(\alpha;\beta)} = \delta^0_\alpha\,\delta^0_\beta\,\dot f - fR\dot R\left(\delta^1_\alpha\,\delta^1_\beta\,\frac{1}{1-kr^2} + \delta^2_\alpha\,\delta^2_\beta r^2 + \delta^3_\alpha\delta^3_\beta r^2\sin^2\theta\right), \tag{2.3}$$

where the dot indicates derivation with respect to t. Calculating $\xi_{(\alpha;\beta);\gamma}$ and equating to zero we obtain the following conditions over the functions $R(t)$ and $f(t)$:

$$f\dot R - R\dot f = 0 \quad \text{and} \quad \ddot f = 0. \tag{2.4}$$

Equations (2.4) immediately give

$$f(t) = cR(t) \quad \text{and} \quad R(t) = at + b, \tag{2.5}$$

where a, b, and c are constants. From (2.5) and (2.3) we see that ξ_α is a homothetic motion of the Robertson–Walker metric when $R(t)$ is a linear function of time.

The Einstein equations for the metric (2.1) with the energy-momentum tensor of a perfect fluid of density ρ and pressure p are

$$\begin{aligned} &2(\ddot R/R) + (\dot R^2 + k)/R^2 = -\kappa p,\\ &3(\dot R^2 + k)/R^2 = \kappa\rho. \end{aligned} \tag{2.6}$$

Equations (2.5) and (2.6) together give

$$\kappa\rho = -3\kappa p = (3(a^2+k)/(at+b)^2). \tag{2.7}$$

The fluid given by (2.7) will satisfy the weak, strong, and dominant energy conditions of Hawking–Ellis[4] if

$$a^2 + k \geqslant 0. \tag{2.8}$$

It is interesting to note that ξ_α of Eq. (2.2) with conditions (2.4) is a homothetic motion of any metric of the form

$$ds^2 = dt^2 - R^2(t)h_{ij}(x^1,x^2,x^3)\,dx^i\,dx^j, \quad i,j = 1,2,3. \tag{2.9}$$

[a)] On leave of absence from Instituto de Física, Universidade Federal do Río de Janeiro, Río de Janeiro, Brazil.

 0022-2488/86/092360-02$02.50

III. CONSTANTS OF MOTION AND GEODESICS

A constant of motion for a free-falling particle can be constructed with affine collineations[2]

$$K = P^\alpha \xi_\alpha - (s/M) P^\alpha \xi_{\alpha;\beta} P^\beta, \tag{3.1}$$

where $P^\alpha = M(dx^\alpha/ds)$ (M is the mass of the particle; s is the proper time). In the case of our collineation (with $a = 1$, $b = 0$, and $c = 1$)

$$\xi_\alpha = t\delta^0_\alpha, \tag{3.2}$$

this constant is

$$K = M\left(t\frac{dt}{ds} - s\right). \tag{3.3}$$

Equation (3.3) can be integrated to give

$$t^2 - s^2 - 2(K/M)s + 2\widetilde{K} = 0, \tag{3.4}$$

where $\widetilde{K}$ is a constant of integration. Now using the first integral of motion

$$g_{\alpha\beta}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} = 1$$

and choosing

$$\frac{d\theta}{ds} = \frac{d\varphi}{ds} = 0$$

(which is possible due to conservation of angular momentum) we have

$$\left(\frac{dt}{ds}\right)^2 - \frac{t^2}{1-kr^2}\left(\frac{dr}{ds}\right)^2 = 1. \tag{3.5}$$

Making a change of variables

$$\begin{aligned} r &= \sin\alpha, \quad \text{for } k = +1,\\ r &= \sinh\alpha, \quad \text{for } k = -1, \\ r &= \alpha, \quad \text{for } k = 0, \end{aligned} \tag{3.6}$$

and using (3.4) we obtain

$$\alpha = \ln\left[\frac{A}{t}\left(\sqrt{t^2 + \frac{K^2}{M^2} + 2\widetilde{K}} - \sqrt{\frac{K^2}{M^2} + 2\widetilde{K}}\right)\right], \tag{3.7}$$

which describes geodesics of the Robertson–Walker spacetime. Due to homogeneity and isotropy of this metric, these geodesics can be used to obtain arbitrary timelike geodesics. The null geodesics are obtained starting from the conservation law

$$K = \frac{dx^\alpha}{d\tau}\cdot\xi_\alpha, \tag{3.8}$$

which yields

$$\alpha = \ln At. \tag{3.9}$$

Equations (3.7) and (3.9) are special cases of geodesics given by Tolman.[5]

ACKNOWLEDGMENTS

We thank Dr. S. Hojman for interesting discussion and criticism. We also thank the Centro de Estudios Nucleares and especially Dr. M. P. Ryan, Jr. for kind hospitality.

[1] G. H. Katzin, J. Levine, and W. R. Davis, J. Math. Phys. **10**, 617 (1969).
[2] S. Hojman, L. Nuñez, A. Patiño, and H. Rago, J. Math. Phys. **27**, 281 (1986).
[3] G. H. Katzin and J. Levine, J. Math. Phys. **9**, 8 (1968).
[4] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U.P., London, 1973).
[5] R. C. Tolman, *Relativity Thermodynamics and Cosmology* (Clarendon, Oxford, 1934).

Gravitational solutions, including radiation, for a perturbed light beam

Raymond W. Nackoney
Department of Natural Science, Loyola University of Chicago, Chicago, Illinois 60626

(Received 1 August 1985; accepted for publication 30 April 1986)

Linearized field equations and solutions are derived for a perturbed sheet beam of light. The work is based on an exact solution of a collimated beam in the geometrical limit. The linearized field changes of the initially curved background metric can be put, with the help of the harmonic conditions, into a normal coordinate form. These six normal coordinates satisfy six linearized, inhomogeneous, field equations in three variables. Stationary solutions include divergent beams. Gravitational waves propagating opposite to the beam's flux are found to be confined to a region about the propagation axis of the beam, much as is experienced in wave guides. Radiative cases can be produced by large angle scattering of light and are discussed in terms of their effect on an ideal optical antenna. The effect is one that grows linearly with time. The growth time is prohibitively long for the most energetic systems that can be realistically considered in the foreseeable future.

I. INTRODUCTION

The gravitational field of a beam of light has been of interest at least as far back as 1922 when Lorentz[1] suggested that two light beams traveling in the same direction may deflect each other due to their gravitational interaction. An approximate weak field solution for finite pencils of light in a vacuum was obtained in 1931 by Tolman, Ehrenfest, and Podolsky.[2,3] Recently Scully[4] extended this work to propagation through a refractive medium. An exact solution for vacuum propagation of a beam or pulse of light was obtained by Bonner[5] in the context of the source of plane gravitational waves, and independently by Nackoney[6,7] as an extension of the work of Tolman *et al.* A further development of Nackoney's work by Banerjee[8] includes an exact solution of the Einstein–Maxwell equations.

The present paper is based on Refs. 6 and 7, which are here referred to as I and II, respectively. In those papers a collimated beam or pulse of unidirectional light of circular cross section and of variable intensity and duration was considered in the geometrical limit far from sources and absorbers of the light. An exact solution resulted. In Sec. II of the present paper the exact solution is presented for a semi-infinite, rectangular cross section; i.e., a sheet of light propagating along one coordinate axis, of finite extent along the second axis, and of infinite extent without variation along the third axis. This choice of cross section permits the use of harmonic conditions. The results of interest are not expected to differ significantly from the circular cross-sectional case.

Section III derives the first-order equations for two classes of perturbations of the initial beam. Essentially, the first is for small angle deflections of the bulk of the beam and the second is for large angle deflections of small portions of the beam. By reduction of the ten linearized field equations for small changes in the metric, six normal coordinates result, satisfying inhomogeneous wave equations in three variables. Section IV considers small angle deflections of the beam and yields nonradiative solutions including the case of a divergent beam. Since the divergent case and the following wave cases are each described by first-order changes in the initial beam, their simultaneous presence will not interact in the first order but will provide cross terms only in the second order. Therefore, restricting ourselves to a first-order calculation allows us to analyze each case separately and superimpose the solutions in dealing with radiative divergent beam cases. The wave solutions of Secs. V and VI have many similarities with the vast body of gravitational wave research. An excellent review of this literature is found in Thorne's article.[9] However, distinctive features do arise, as in Sec. V where homogeneous wave solutions are considered. For example, we find that the beam's gravitational trough acts as a wave guide for gravitational waves propagating opposite to the beam's flux. In addition, the normal coordinates yield two nonzero modes that mimic, but do deviate from, the strict transverse traceless polarization of waves in Minkowski space. Section VI looks at the inhomogeneous equations and yields gravitational radiation from large angle scattering of light. This type of source is one that does not allow a slow motion approximation. Specifically, an analysis is presented of a gravitational radiator created by light oscillating parallel to the beam between two mirrors. This gravitational radiation is guided by the beam back to a second optical oscillator, which acts as an antenna. The resulting interaction is found to change the optical energy in the antenna. This change grows linearly with time. The magnitude of this interaction is quite small and is discussed in some detail.

II. INITIAL COLLIMATED BEAM

Consider a sheet beam of light in a rectangular coordinate system $(t,x,y,z) = (x^0,x^1,x^2,x^3)$. The beam will propagate along the z axis toward increasing z values. It will be considered in the geometrical limit so that the optical wavelength goes to zero and, hence, the initial beam may remain collimated. The cross section of the beam in the x-y plane is chosen to be an infinite rectangular sheet; infinite and unchanging along the y axis and of finite but variable extent $\pm X$ along the x axis. The energy density ρ of the light beam

 0022-2488/86/092362-11$02.50

will be variable and include pulses of light rather than just continuous beams.

The functional dependence of ρ is simplest if we position the absorber and the source of the beam at $\pm$ infinity along the z axis. We introduce a retarded time t so that elements of the light beam in a given x-y cross section are each labeled by a single value of t, regardless of their subsequent z position. Hence ρ will be a function of t, but not of z, for the collimated beam. Variations with x across a given cross section will be allowed. Variations with y will not be considered. Hence $\rho=\rho(t,x)$ and $X=X(t)$. We define a scaled energy density as in Eq. (I, 44) [see paper I, Eq. (44)] and Eq. (II, 6) by

$$m(t,x)=4\pi Gc^{-4}\rho\,. \tag{2.1}$$

Since we are working in the geometrical limit, the energy-momentum tensor is that of a null fluid with the scaled local energy density given by Eq. (2.1). Combining the field analyses leading to Eq. (I, 51) and Eq. (II, 25), and setting $c=1$, we find

$$ds^2=f(t,x)dt^2+2\,dt\,dz-dx^2-dy^2\,, \tag{2.2}$$

where

$$\frac{\partial^2 f}{\partial x^2}-4mf=0\,. \tag{2.3}$$

For each value of t, f satisfies the equation for m at that time. Hence, f follows m in its time dependence. The field travels with and depends only on the cross section of the beam which shares the same t value independent of the z position.

If we choose m independent of x, then the solution to Eq. (2.3) is

$$\begin{aligned} f&=\cosh(2m^{1/2}x)\,,\quad |x|\leqslant X\,,\\ &=A+B\,|x|\,,\quad |x|\geqslant X\,, \end{aligned} \tag{2.4}$$

where

$$A=\cosh(2m^{1/2}X)-2m^{1/2}X\sinh(2m^{1/2}X)$$

and

$$B=2m^{1/2}\sinh(2m^{1/2}X)\,,$$

where m and X may be functions of t. Regardless of the x dependence of m the exterior solution is linear in x with singularities at $\pm$ infinity.

III. PERTURBATION EQUATIONS

A. First-order setup

Two classes of physical changes from the initial (zero-order) beam will be allowed. First, a small change Δu^i in the beam's four-velocity from u^i to $\bar{u}^i=u^i+\Delta u^i$ with the only nonzero component of the initial four-velocity being $u^3=1$. We continue to suppress all changes with y. Hence $\Delta u^2=0$ and the remaining $\Delta u^i=\Delta u^i(t,x,z)$. This change gives rise to a new scaled energy density $\overline{m}=\overline{m}(t,x,z)$, which can differ from $m(t,x)$ by a first-order quantity. The second class of physical changes will be those in which a small portion $\Delta m(t,x,z)$ of the beam's energy density is propagated in a new direction given by the four-velocity $\hat{u}(t,x,z)$. We maintain our suppression of y and set $\hat{u}^2$ to zero.

These two changes give rise to a new metric tensor

$$g_{ik}=g_{(0)ik}+h_{ik}(t,x,z)\,, \tag{3.1}$$

where $g_{(0)ik}$ is the zero-order metric of Eq. (2.2). We consider those changes that result in the gravitational field equations having no more than a linear dependence on the first-order quantities h_{ik}, Δu^i, Δm, and $\overline{m}-m$, where quantities quadratic in these changes are deemed negligible. The null condition on the four-velocity $\bar{u}^i$ gives

$$\Delta u^0=-\tfrac{1}{2}h_{33}\,. \tag{3.2}$$

Therefore, if h_{33} is not equal to zero, dt changes as one proceeds along a null geodesic. For the beam, this means that t would cease to be a retarded time and differ from it by a first-order quantity. However, in dealing with quantities that are themselves first order, this difference would be of second order and thus be negligible. Therefore, the retarded time description holds for all first-order quantities. One significant change is the existence of a z dependence in first-order quantities. This allows for the initiation of a perturbation at a unique position along the z axis, and results in the loss of the equivalence of observers along the beam axis.

If we assume that m is small enough so that the photon–photon interactions of intersecting light rays are negligible, then we effectively have two null fluids that can interpenetrate without interaction. In this case, the energy-momentum tensor to first order is $T_{ik}=\overline{T}_{ik}+\widehat{T}_{ik}$, where, by Eq. (I, 25),

$$\begin{aligned} \overline{T}_{ik}&=\overline{m}g_{00}\bar{u}_i\bar{u}_k/\bar{u}_0^2\\ &\text{and}\\ \widehat{T}_{ik}&=\Delta m g_{00}\hat{u}_i\hat{u}_k/\hat{u}_0^2\,. \end{aligned} \tag{3.3}$$

Conditions on T_{ik} exist. First, since we are dealing with null fluids, the trace $T^i{}_i=0$ and this gives

$$2\widehat{T}_{03}-\widehat{T}_{11}-f\widehat{T}_{33}=0\,. \tag{3.4}$$

Second, the conservation equations, $T_i{}^k{}_{;k}=0$, (in the sequence $i=0,1,2,3$) give

$$\begin{aligned} &m\frac{\partial h_{00}}{\partial z}-\frac{1}{2}mf\frac{\partial h_{11}}{\partial z}-\frac{1}{2}mf\frac{\partial h_{22}}{\partial z}-mf\frac{\partial h_{33}}{\partial t}\\ &\quad-\frac{1}{2}m\frac{\partial f}{\partial t}h_{33}-\frac{1}{2}\frac{\partial m}{\partial t}fh_{33}+f\frac{\partial\overline{m}}{\partial z}+\frac{\partial(mf\Delta u')}{\partial x}\\ &\quad+\frac{\partial\widehat{T}_{00}}{\partial z}-\frac{\partial\widehat{T}_{01}}{\partial x}+\frac{\partial\widehat{T}_{03}}{\partial t}-f\frac{\partial\widehat{T}_{03}}{\partial z}-\frac{1}{2}\frac{\partial f}{\partial t}\widehat{T}_{33}=0\,, \end{aligned} \tag{3.5}$$

$$\begin{aligned} &mf\frac{\partial h_{13}}{\partial z}-\frac{1}{2}mf\frac{\partial h_{33}}{\partial x}-mf\frac{\partial\Delta u'}{\partial z}+\frac{\partial\widehat{T}_{01}}{\partial z}\\ &\quad-\frac{\partial\widehat{T}_{11}}{\partial x}+\frac{\partial\widehat{T}_{13}}{\partial t}-f\frac{\partial\widehat{T}_{13}}{\partial z}-\frac{1}{2}\frac{\partial f}{\partial x}\widehat{T}_{33}=0\,, \end{aligned} \tag{3.6}$$

$$mf\frac{\partial h_{23}}{\partial z}=0\,, \tag{3.7}$$

and

$$\frac{\partial\widehat{T}_{03}}{\partial z}-\frac{\partial\widehat{T}_{13}}{\partial x}+\frac{\partial\widehat{T}_{33}}{\partial t}-f\frac{\partial\widehat{T}_{33}}{\partial z}=0\,. \tag{3.8}$$

B. Linearized field equations

With the current scale factors, the field equations, with the cosmological constant set to zero, are $R_{ik}=-2T_{ik}$. We find ten linear, coupled, inhomogeneous, second-order, differential equations in three variables. The ten equations naturally separate into a set of three equations for three of the h_{ik}—h_{02}, h_{12}, h_{23}—and a set of seven equations for the remaining seven h_{ik}. The situation is similar to the analysis of a system of coupled oscillators. In that problem the isolation of normal coordinates which undergo simple harmonic motion is advantageous. A similar procedure is possible here.

We first impose harmonic conditions $\Gamma^k \equiv g^{mn}\Gamma^k_{mn} = 0$, and we have (in the sequence $k = 0,1,2,3$)

$$\frac{1}{2}\frac{\partial h_{11}}{\partial z} - \frac{\partial h_{13}}{\partial x} + \frac{1}{2}\frac{\partial h_{22}}{\partial z} + \frac{\partial h_{33}}{\partial t} - \frac{1}{2}f\frac{\partial h_{33}}{\partial z} = 0\,, \tag{3.9}$$

$$\frac{\partial h_{01}}{\partial z} - \frac{\partial h_{03}}{\partial x} - \frac{1}{2}\frac{\partial h_{11}}{\partial x} + \frac{\partial h_{13}}{\partial t} - f\frac{\partial h_{13}}{\partial z} + \frac{1}{2}\frac{\partial h_{22}}{\partial x} + \frac{1}{2}f\frac{\partial h_{33}}{\partial x} + \frac{1}{2}\frac{\partial f}{\partial x}h_{33} = 0\,, \tag{3.10}$$

$$\frac{\partial h_{02}}{\partial z} - \frac{\partial h_{12}}{\partial x} + \frac{\partial h_{23}}{\partial t} - f\frac{\partial h_{23}}{\partial z} = 0\,, \tag{3.11}$$

and

$$\frac{\partial h_{00}}{\partial z} - \frac{\partial h_{01}}{\partial x} - f\frac{\partial h_{03}}{\partial z} + \frac{1}{2}\frac{\partial h_{11}}{\partial t} + \frac{\partial f}{\partial x}h_{13} + \frac{1}{2}\frac{\partial h_{22}}{\partial t} + \frac{1}{2}f\frac{\partial}{\partial t}\frac{\partial h_{33}}{\partial t} - \frac{1}{2}\frac{\partial f}{\partial t}h_{33} = 0\,. \tag{3.12}$$

By taking first partials of the harmonic conditions and forming suitable linear combinations of these, the undifferentiated conditions times $\partial f/\partial t$, and Eq. (2.3) we may put the field equations into a new form.[10] In this form the differential operator on the h_{ik} is exclusively of d'Alembertian form $\Box$, where

$$\Box \equiv -g_{(0)}{}^{ik}\,\partial_i\,\partial_k = \frac{\partial^2}{\partial x^2} - \frac{2\partial^2}{\partial t\,\partial z} + f\frac{\partial^2}{\partial z^2}\,. \tag{3.13}$$

The resulting d'Alembertian form of the field equations is

$$\Box h_{23} = 0\,, \tag{3.14}$$

$$\Box h_{12} = -\frac{\partial f}{\partial x}\frac{\partial h_{23}}{\partial z}\,, \tag{3.15}$$

$$\Box h_{02} = -\frac{\partial f}{\partial x}\frac{\partial h_{12}}{\partial z} + \frac{\partial f}{\partial x}\frac{\partial h_{23}}{\partial x} - \frac{\partial f}{\partial t}\frac{\partial h_{23}}{\partial z} + 4mf\,h_{23}\,, \tag{3.16}$$

$$\Box h_{22} = 0\,, \tag{3.17}$$

$$\Box h_{33} = 4\hat{T}_{33}\,, \tag{3.18}$$

$$\Box h_{13} = -\frac{1}{2}\frac{\partial f}{\partial x}\frac{\partial h_{33}}{\partial z} + 4\hat{T}_{13}\,, \tag{3.19}$$

$$\Box h_{03} = \frac{\partial f}{\partial x}\frac{\partial h_{33}}{\partial x} - \frac{3}{2}\frac{\partial f}{\partial t}\frac{\partial h_{33}}{\partial z} + 2mf\,h_{33} + 4\hat{T}_{03}\,, \tag{3.20}$$

$$\Box h_{11} = -2\frac{\partial f}{\partial x}\frac{\partial h_{13}}{\partial z} + \frac{\partial f}{\partial x}\frac{\partial h_{33}}{\partial x} + 4\hat{T}_{11}\,, \tag{3.21}$$

$$\Box h_{01} = -\frac{\partial f}{\partial x}\frac{\partial h_{03}}{\partial z} - \frac{1}{2}\frac{\partial f}{\partial x}\frac{\partial h_{11}}{\partial z} + \frac{\partial f}{\partial x}\frac{\partial h_{13}}{\partial x} - \frac{\partial f}{\partial t}\frac{\partial h_{13}}{\partial z} + 4mf\,h_{13} + \frac{1}{2}\frac{\partial f}{\partial x}\frac{\partial h_{22}}{\partial z} - \frac{1}{2}\frac{\partial f}{\partial t}\frac{\partial h_{33}}{\partial x} + \frac{3}{2}\frac{\partial f}{\partial x}\frac{\partial h_{33}}{\partial t} - \frac{1}{2}f\frac{\partial f}{\partial x}\frac{\partial h_{33}}{\partial z} - 4mf\Delta u' + 4\hat{T}_{01}\,, \tag{3.22}$$

and

$$\Box h_{00} - 4mh_{00} = -2\frac{\partial f}{\partial x}\frac{\partial h_{01}}{\partial z} + 2\frac{\partial f}{\partial x}\frac{\partial h_{03}}{\partial x} - 2\frac{\partial f}{\partial t}\frac{\partial h_{03}}{\partial z} - \frac{\partial^2 f}{\partial x^2}h_{11} + 2\frac{\partial f}{\partial x}\frac{\partial h_{13}}{\partial t} + 2\frac{\partial^2 f}{\partial t\,\partial x}h_{13} - \frac{\partial f}{\partial t}\frac{\partial h_{33}}{\partial t} - \frac{\partial^2 f}{\partial t^2}h_{33} - \left(\frac{\partial f}{\partial x}\right)^2 h_{33} + 4(\bar{m} - m)f + 4\hat{T}_{00}\,. \tag{3.23}$$

We now introduce six normal coordinates H_{ik}. Three are obvious and the remaining three are patterned after the harmonic conditions. The two normal coordinates for the set of three equations are

$$H_{23} = h_{33} \tag{3.24}$$

and

$$\frac{\partial H_{12}}{\partial z} = \frac{\partial h_{12}}{\partial z} - \frac{\partial h_{23}}{\partial x}\,. \tag{3.25}$$

The four for the set of seven equations are

$$H_{22} = h_{22}\,, \tag{3.26}$$

$$H_{33} = h_{33}\,, \tag{3.27}$$

$$\frac{\partial H_{13}}{\partial z} = \frac{\partial h_{13}}{\partial z} - \frac{1}{2}\frac{\partial h_{33}}{\partial x}\,, \tag{3.28}$$

and

$$\frac{\partial H_{03}}{\partial z} = \frac{\partial h_{03}}{\partial z} - \frac{1}{2}\frac{\partial h_{33}}{\partial t} - \frac{1}{2}f\frac{\partial h_{33}}{\partial z}\,. \tag{3.29}$$

By using the field equations, the harmonic conditions, the conditions on the energy-momentum tensor, and the null condition [Eq. (3.2)] it can be shown that the normal coordinates satisfy the following:

$$\Box H_{12} = 0 , \tag{3.30}$$

$$\Box H_{23} = 0 , \tag{3.31}$$

$$\Box H_{22} = 0 , \tag{3.32}$$

$$\Box H_{33} = 4\hat{T}_{33} , \tag{3.33}$$

$$\Box H_{13} = 4\hat{T}_{13} - 8 \int \frac{\partial \hat{T}_{33}}{\partial x} dz , \tag{3.34}$$

and

$$\Box H_{03} = 4\hat{T}_{03} - 2f\hat{T}_{33} - 2 \int \frac{\partial \hat{T}_{33}}{\partial t} dz . \tag{3.35}$$

Once the above are solved, the solution for the ten h_{ik} can be obtained from the definitions of the six normal coordinates Eqs. (3.24)–(3.29) and the four harmonic conditions Eqs. (3.9)–(3.12).

IV. SMALL ANGLE DEFLECTIONS

A. General equations

Realistic beams of light cannot remain collimated, but undergo a diffractive divergence. This is a case which fits the first class of perturbations. In general, to analyze this case we set Δm and $\hat{u}^i$ to zero. Hence $\hat{T}_{ik}$ is zero and Eqs. (3.30)–(3.35) for the normal coordinates become homogeneous. Now H_{03}, H_{13}, H_{23}, and H_{33} can be set to zero by the remaining freedom in the coordinates. The homogeneous equations for H_{12} and H_{22} are without boundary conditions which relate specifically to this class of problems. Hence, we may set these to zero in this section. The harmonic conditions [Eqs. (3.9)–(3.12)] determine the h_{ik} to be polynomials in z of up to cubic terms in z for h_{00}. The polynomial coefficients are interrelated functions of t and x. Working with the d'Alembertian form of the field equations and demanding continuity of the metric up to its first derivatives yields two nonzero h_{ik}:

$$h_{01} = \zeta(t,x) \tag{4.1}$$

and

$$h_{00} = \frac{\partial \zeta}{\partial x} (z - z_0) + \eta(t,x) , \tag{4.2}$$

where the constant z_0 and the functions ζ and η are determined by the perturbation. The conservation Eq. (3.6) yields $\Delta u^1 = \Delta u^1(t,x)$ and this in the conservation equation (3.5) gives

$$\overline{m} = m(t,x) + \mu_0(t,x) + \mu_1(t,x)(z - z_0) , \tag{4.3}$$

where

$$\mu_1 = - \frac{\partial m \Delta u^1}{\partial x} - \frac{m}{f} \int_{x_0(t)}^{x} \frac{\partial f}{\partial x'} \frac{\partial \Delta u^1}{\partial x'} dx'$$

and μ_0 is a first-order function determined by the specific perturbation. These, in conjunction with the field equations (3.22) and (3.23), yield

$$\frac{\partial^2 \zeta}{\partial x^2} = - 4mf\Delta u^1 \tag{4.4}$$

and

$$\frac{\partial^2 \eta}{\partial x^2} - 4m\eta = 2 \frac{\partial^2 \zeta}{\partial t \, \partial x} + 4f\mu_0 . \tag{4.5}$$

Since $h_{33} = 0$, Eq. (3.2) yields $\Delta u^0 = 0$ and the retarded time description of the beam's cross sections is maintained. If we consider the geodesic equations we may show that within the region where this class of perturbations exists, $\Gamma^i_{33} = 0$ and hence $d\bar{u}^i/ds = 0$. Thus, the four-velocity of the beam's elements does not change along its geodesic under this type of perturbation, and a diverging beam continues to diverge at a constant angle.

B. Divergent beams

The specific case that we will consider is a beam divergent in a region starting at z_0 along the z axis. We choose the simple linearly spreading case in which m and X are initially constant and the first-order change is given by

$$\Delta u^1 = \sigma x , \tag{4.6}$$

where σ is a constant first-order quantity that parametrizes the rate of divergence. Under this change a beam element initially at x at z_0 propagates to $x(1 + \sigma(z - z_0))$ at z. This includes the boundary rays at $\pm X$. The energy density also changes, by Eq. (4.3),

$$\overline{m} = m[1 - \sigma((2f + 1)/f)(z - z_0)] . \tag{4.7}$$

Since we demand continuity of g_{01}, g_{00}, and their first derivatives at $x = 0$ and at the displaced beam boundaries, we find in the divergent region that η is identically zero and

$$\begin{aligned} \zeta &= \sigma F_1(x) , \quad |x| < X , \\ &= \sigma[F_1(X) + (x - X)F_1'(X)] , \quad x > X , \\ &= \sigma[-F_1(X) + (x + X)F_1'(X)] , \quad x < -X , \end{aligned} \tag{4.8}$$

where

$$F_1(x) = 2\alpha^{-1} \sinh(\alpha x) - x - x \cosh(\alpha x)$$

and $\alpha = 2\sqrt{m}$. To visualize this solution we may rewrite Eq. (4.7). Since $\alpha(z - z_0)$ is the fractional change $\Delta x/x$ in a ray's axial displacement we have

$$\overline{m}/m = (1 - \Delta x/x) - (1 - f^{-1})\Delta x/x .$$

The first term on the right is the classical divergence. The second term [since $f(0) = 1$] is a purely relativistic term and corresponds to a gravitational red shift of the energy density with one factor of $f^{-1/2}$ from the energy change of a light ray and a second $f^{-1/2}$ from the invariant scalar density $\sqrt{-g}\, d\Omega$ of the four-volume element.

In paper I we found that other than beam rays, null geodesics in the x-z plane generally propagate toward decreasing z values in a series of meanders of lobe-shaped loops that have a constant maximum displacement $\pm \tilde{x}$ from the beam axis. (The word "generally" excludes superenergetic beams that could generate a Schwarzschild singularity if absorbed within a length and time corresponding to a beam diameter.) The changes in the geodesics due to a divergent beam can be easily investigated external to the beam. By Eqs. (4.1), (4.2), and (4.8), $h_{01} = a + bx$ and $h_{00} = bz$, where a and b are first-order constants with $b < 0$. Two geodesic equations are worth considering. First, $du^0/ds = \frac{1}{2}bu^0u^0$. The solution to first order for a ray moving in the x-z plane is $u^0 = A(1 + \frac{1}{2}b(t - t_0))$. An analysis inside the beam has much the same result with b still negative, but variable and tending to zero at the x origin. Hence, a ray proceeding toward negative z values will undergo a decrease in u^0 as it proceeds into a region in which the beam diverges. Effectively this means a decrease in u^0 during the passage. The second geodesic equation $du^1/ds = -\frac{1}{2}(df/ds)u^0u^0$ is formally unchanged, but is now difficult to solve since u^0 is changing. We may get a qualitative solution using the unperturbed solution $u^1 = \pm(2A - A^2f)^{1/2}$ with the recognition that $u^0 = A$ decreases with time. The maximum displacement $\tilde{x}$ of a null geodesic in the x-z plane is given by the above to be $f(\tilde{x}) = 2/A$. Since f is a monotonically increasing function with $|x|$ we see that $\tilde{x}$ increases as a ray progresses into the divergent region. Hence, the null geodesics in the x-z plane will still meander about the beam axis as they progress back down the beam, but their envelope will grow. This effect will be of interest in considering gravitational waves impinging on a diverging beam of light.

V. SOURCE-FREE WAVE SOLUTIONS

Wave solutions, including radiative ones, can be obtained in this idealized system of a collimated beam of light. Their application to physically realistic divergent beams is, however, not difficult. Section IV described a divergent beam by the first class of perturbations with a change Δu^i in the beam four-velocity. The present and following sections deal with wave and radiative solutions that can be described by the second class of perturbations in which Δu^1 is zero. Both types of perturbations initiate first-order changes in the field equations, which add, but do not mix, in the first-order analysis. This allows us to ignore a realistic beam's divergence in discussing wave solutions and, if need be, to add it in later.

A. Homogeneous wave solutions

We choose Δm, $\hat{u}^i$, and Δu^i to be zero and $\overline{m} = m$. Equations (3.30)–(3.35) for all six H_{ik} are without a source and become homogeneous. As in Sec. IV, four of the six H_{ik} can be transformed to zero by an appropriate choice of the remaining freedom in our coordinate system. Two normal coordinates remain, satisfying $\Box H_{ik} = 0$. Empty space solutions exist and are equivalent to the well-known solution of linearized, plane, transverse, gravitational waves in Minkowski space-time with a pseudotrace $g^{ik}h_{ik} = h_{11} + h_{22}$, which is zero.

Considering the solutions in the presence of a beam of light, we note that our differential operator $\Box$ has a variable coefficient f as compared to the vacuum case of the previous paragraph. If we consider only those cases in which the initial beam has a temporally constant flux, then both m and f are independent of t. Since the coefficients of the differentials of t and z are now independent of t and z, the elementary periodic solutions of the homogeneous equations are of the form

$$\Psi(t,x,z) = \psi(x)\exp[\pm i(\omega t + \kappa z)]\,, \tag{5.1}$$

where ω and κ are real constants and ψ satisfies

$$\frac{d^2\psi}{dx^2} + (2\omega\kappa - \kappa^2 f(x))\psi = 0\,. \tag{5.2}$$

The function ψ is subject to the boundary conditions that it is bounded and should go to zero as x goes to $\pm$ infinity. In addition, ψ and its first derivative must be continuous. With $f(x) \geqslant 1$ and symmetric about $x = 0$, only trivial solutions exist for ψ for $2\omega/\kappa \leqslant 1$. Therefore, we choose $2\omega/\kappa > 1$. Since f is a monotonically increasing function of $|x|$, the coefficient of the linear term in ψ in Eq. (5.2) changes sign at some value of x, say $\tilde{x}$. That value is given by

$$f(\pm\tilde{x}) = 2\omega/\kappa\,. \tag{5.3}$$

Parametrizing Eq. (5.2) by ω and $\tilde{x}$, rather than by ω and κ, we have

$$\frac{d^2\psi}{dx^2} + \frac{4\omega^2}{f(\tilde{x})^2}[f(\tilde{x}) - f(x)]\psi = 0\,. \tag{5.4}$$

For $|x| < \tilde{x}$ the solution ψ will be oscillatory in x. For $|x| > \tilde{x}$, $|\psi|$ will decrease monotonically to zero. The decrease is at a faster rate than a pure exponential since f increases with $|x|$. Because of this, ψ will be effectively concentrated in the region $|x| < \tilde{x}$ and we will refer to $\tilde{x}$ as the confinement distance of the solution. The speed of the Ψ wave along the beam axis is given by Eqs. (5.1) and (5.3) to be

$$\frac{dz}{dt} = -\frac{f(\tilde{x})}{2}\,. \tag{5.5}$$

Since f is positive, we find that the phase velocity along the z axis is always negative. This exclusive "backward" propagation of the waves is a result of the nature of the beam and the retarded time metric. A wave propagating forward, i.e., to larger z values, would move with the beam and share its time dependence. Hence, such a wave front would be parametrized by a single value of t independent of z and is related to the acceleration fields of paper II, Sec. 4.

In effect then, the beam's gravitational field acts as a wave guide for gravitational waves propagating opposite to the beam's flux. Since f varies with x, the strongest analogy would be to an optical fiber with a variable index of refraction. Ray tracing techniques applicable for optical fibers are comparable to the analysis of null geodesics of the unperturbed beam. As mentioned in Sec. IV B, the null geodesics in the x-z plane are a set of meanders, or lobe-shaped loops, which alternate about the z axis and generally progress toward negative z values. We will associate the null geodesic with a maximum x displacement at $\pm\tilde{x}$ with the solution that has its confinement distance to be $\tilde{x}$. We may calculate dz/dt for this geodesic at its maximum displacement $\tilde{x}$ from the line element Eq. (2.2). This speed matches the value given by Eq. (5.5) for the phase velocity of the wave. Hence, there is a correlation between our solution and a specific null geodesic, or better, a family of null geodesics that differ only by a displacement along the z axis. This entire family can be said to characterize the wave solution. Since the average speed along the z axis of the geodesic is less than that of Eq. (5.5), the geodesic that represents a given wave front at a specific time is replaced in the next moment by a neighboring geodesic in the family. This sequential representation will be convenient in Sec. V.

With the boundary conditions cited above we should expect that for each confinement distance there will be a discrete sequence of ω; namely, the characteristic values ω_n

where n ranges through the non-negative integers. Our solutions to Eq. (5.4) can be written as $\psi_{\tilde{x},n}(x)$. This provides a complete set of characteristic functions for each value of $\tilde{x}$. The orthogonality condition for the set is

$$\int_{-\infty}^{\infty} \frac{f(\tilde{x}) - f(x)}{f(\tilde{x})^2} \psi_{\tilde{x},n}(x)\psi_{\tilde{x},m}(x)\, dx = M_n(\tilde{x})\delta_{mn} . \tag{5.6}$$

It may be noted that of all the variety of ways in which ω and κ could be stipulated, only fixing ω and the ratio ω/κ in the manner chosen yields a suitable orthogonality relation.

The exterior $|x| > X$ solution of Eq. (5.4) can be written in terms of Bessel functions since f is given by Eq. (2.4) and is linear in this region. Equation (5.4) becomes in this region

$$\frac{d^2\psi}{dx^2} + \frac{4\omega_n^2 B}{(A + B\tilde{x})^2}(\tilde{x} - x)\psi = 0 .$$

For $X < |x| < \tilde{x}$ the solutions are in terms of $J_{1/3}$ and $J_{-1/3}$. For $|x| > \tilde{x}$ the boundary conditions at infinity demand that the modified Bessel functions $K_{1/3}$ be used, e.g., for $|x| > X$ and $|x| > \tilde{x}$:

$$\psi = a(x - \tilde{x})^{1/2} K_{1/3}\left[\frac{4\omega B^{1/2}}{3(A + B\tilde{x})}(x - \tilde{x})^{3/2}\right], \tag{5.7}$$

where a is a matching constant.

The interior solution can be solved for the case in which the scaled energy density m is a constant. In this case Eq. (2.4) in Eq. (5.4) yields a Mathieu equation. Since any conceivable physical beam has a very small value of mX^2 [see Eq. (5.10)] we approximate

$$f \approx 1 + 2mx^2, \quad x < X . \tag{5.8}$$

With this, Eq. (5.4) can be put into the form of Weber's differential equation

$$\frac{d^2\psi}{d\zeta^2} + \left[\nu + \frac{1}{2} - \frac{1}{4}\zeta^2\right]\psi = 0 ,$$

with

$$\nu = \omega(2m)^{1/2}\tilde{x}^2/(1 + 2m\tilde{x}^2) - \tfrac{1}{2}$$

and

$$\zeta = [4\omega(2m)^{1/2}/(1 + 2m\tilde{x}^2)]^{1/2} x .$$

The solutions are the even and odd parabolic cylinder functions $E_\nu^{(0)}(\zeta)$ and $E_\nu^{(1)}(\zeta)$, respectively. These functions are given by

$$E_\nu^{(0)}(\zeta) = \sqrt{2}\, e^{-\zeta^2/4}\, {}_1F_1(-\nu/2;\tfrac{1}{2};\zeta^2/2)$$

and (5.9)

$$E_\nu^{(1)}(\zeta) = 2\zeta e^{-\zeta^2/4}\, {}_1F_1((1 - \nu)/2;\tfrac{3}{2};\zeta^2/2) ,$$

where ${}_1F_1$ is the confluent hypergeometric function

$${}_1F_1(\alpha;\beta;z) = \frac{\Gamma(\beta)}{\Gamma(\alpha)} \sum_{\lambda=0} \frac{\Gamma(\alpha + \lambda)z^2}{\Gamma(\beta + \lambda)\lambda!} .$$

By alternately choosing $E^{(0)}$ or $E^{(1)}$ as the interior solutions for ψ we can produce a symmetrical positioning of nodal points. For a specific beam with set m and X, our matched solutions $\psi_{\tilde{x},n}$ will range over integer $n \geqslant 0$ for each confinement distance $\tilde{x}$. With this choice n will be the number of nodal points of the solution and $\psi_{\tilde{x},n}$ is an even or odd function dependent on the evenness or oddness of n. By choosing $\tilde{x} \leqslant X$ we simplify the matching of the interior solutions Eqs. (5.9) with the exterior solutions Eqs. (5.7) without qualitatively changing the analysis. We may write

$$mX^2 = (\pi G/c^5)P = 8.66 \times 10^{-53} P , \tag{5.10}$$

where P is the power of the sheet beam in watts per length $2X$ in the y direction (per square cross section). Because of the smallness of the coefficient in Eq. (5.10) we find that the sequence of characteristic values for physically plausible powers is

$$\omega_n = 1.14 \times 10^{34}(\beta_n/XP^{1/2})(X/\tilde{x})^2 \text{ Hz} , \tag{5.11}$$

where X is in meters and $\tilde{x} \leqslant X$. For $\tilde{x} = X$, $\beta_0 = 0.96$, $\beta_1 = 2.98$, $\beta_2 = 4.99$, and, as n grows, $\beta_n \to 2n + 1$. For $\tilde{x} \ll X$ the fit improves and β_n is closely approximated by $2n + 1$ even for small n. For example, at $\tilde{x} = 0.5X$, $\beta_0 = 1.00, \beta_1 = 3.00$, etc. The $2n + 1$ linearity in characteristic values is similar to the energy eigenvalue spectrum of the quantum simple harmonic oscillator. The similarity includes $\tilde{x}$, which becomes the edge of the classically allowed region of the oscillator. The main difference in the two problems is that in our case the potential well of our gravitational trough becomes linear for $x > X$. However, for $\tilde{x} \leqslant X$ the $\psi_{\tilde{x},n}$ solution is decreasing rapidly for $x > X$. In this situation, especially for large n, the amount of $\psi_{\tilde{x},n}$ that penetrates beyond the beam boundary X is relatively small. Hence, $\psi_{\tilde{x},n}$ is not sensitive to the exact form of the potential well for $x > X$. The smaller $\tilde{x}$ is with respect to X the less important the external solution of f becomes, and β_n approaches $2n + 1$ for all n.

With the identification with the well-known quantum simple harmonic oscillator, we may condense our discussion of $\psi_{\tilde{x},n}(x)$. As with the quantum case, $\psi_{\tilde{x},n}$ becomes enhanced just interior to $\tilde{x}$ as the value of n grows. In the classical oscillator this corresponds to the relatively slow speed of, say, a pendulum bob near its maximum amplitude. In our case we may look at the null rays that characterize our solution. These rays snake down the gravitational trough of the light beam, opposite to the beam's flux, with a maximum displacement at $\pm\tilde{x}$. At either of these extremes the ray is propagating parallel to the beam axis. Hence it will spend much of its "time" near these extremes and $\psi_{\tilde{x},n}$ will be enhanced there.

For $\tilde{x} > X$ we qualitatively have the same form of the solution. The one additional feature worth noting is that as $\tilde{x}$ increases, the ω_n values become a continuum. In so doing, $\psi_{\tilde{x},n}$ represents very x-extended gravitational waves propagating opposite to the beam which, in the limit, become plane waves.

B. Effect of the source-free wave solutions

As with the divergent beam solution, $\Gamma^i_{33} = 0$ and $du^i/ds = 0$. The beam geodesics are unaffected to first order by the passage of the homogeneous wave and the beam remains collimated. Oscillations in the beam's energy density do occur and can be calculated by the conservation Eq. (3.5). We find that $\partial(g_{00}\overline{m})/\partial z = 0$. Since g_{00} undergoes a first-order change with z, so will the local energy density $\overline{m}$.

If we assume that we can extend this solution of a beam in the geometrical limit to a light beam with a finite frequency, then we have the possibility of a heterodyne effect with the gravitational wave. Considering that $\psi_{\tilde{x},n}$ depends on x, then the heterodyning would have an x dependence acting as the signature of a gravitational wave's passage. However, the effect is of first order and we must contend with the gravitational wave frequency as given by Eq. (5.11). If we could use the reported power of a 10^{14} W pulsed x-ray laser[11] and choose $\tilde{x} = X = 1$ m, we find that the fundamental gravitational wave frequency is 10^{27} Hz. Since the x rays are reported to have a frequency of 10^{17} Hz we cannot expect effective heterodyning of the two waves. Although gravitational waves of less than the first resonant frequency might be generated and directed back down the beam, they would have to be classified as highly subresonant. The response of any system to weak impulses of highly subresonant frequencies is always small and in our case would be totally undetectable in the foreseeable future.

An astrophysical case is worth considering. By aiming a laser at a rotating neutron star one might expect that a portion of the gravitational radiation produced by the star would be captured in the beam's gravitational wave guide and either be directed back toward a detector on the Earth or possibly be focused into the beam so that its effect would be amplified. Taking ω_0 to be the rotational frequency of a neutron star, say 10^3 Hz, we may use the solutions when $\tilde{x} > X$. Since the beam would have to be continuous or continually pulsed we choose $P = 10^6$ W. With $X = 1$ m it can be shown that $\tilde{x} \approx 4\times 10^{18}$ m = 400 light years. With x so large, the wave would be greatly spread out and the effect on the beam or detector would be without the distinctive x-dependent signature of a wave where $\tilde{x} \approx X$. In addition, the diffractive divergence of a physical beam will act to spread the gravitational wave even more. This is a result of the conclusion discussed at the end of Sec. IV that the maximum displacement $\tilde{x}$ of the null rays in the x-z plane grows as the rays progress opposite to a divergent beam's flux. Since we have identified a wave with a confinement distance $\tilde{x}$ with these null rays, we may conclude that the wave guide effect is diminished in this case and that the gravitational wave will also diverge upon encountering a divergent beam of light. Once again the effect of such an encounter is not expected to be observable.

VI. RADIATIVE SOLUTIONS

We will describe the solution and effect of a small portion of light caught between two fixed ideal mirrors at z_1 and z_2, respectively, with $z_1 < z_2$. The mirrors will be oriented such that the oscillating light of density Δm oscillates parallel to the beam at an average distance $x = \xi$ from the beam axis, with an x extent about ξ of Δx. So that the mirrors do not interfere with the beam itself we may place them just outside of the beam at $\xi > X$. We will ignore the gravitational field of the mirrors and analyze the field produced by the light alone. With this assumption the analysis can be divided into two. First, we consider the field with the light source moving forward with the beam and, second, we consider the field with the light moving rearward, opposite to the beam's flux. In both cases we will concentrate on the solution in the region $z < z_1$.

A. Forward moving phase of oscillator

In the z_1 to z_2 region the four-velocity u^i of the forward phase of the oscillating light is identical to that of the initial beam. We find only one first-order component of the energy-momentum tensor $\hat{T}_{00} = \Delta m f$. With the d'Alembertian form of the field equations (3.14)–(3.23), only Eq. (3.23) for h_{00} has an inhomogeneous term. The other h_{ik} are without sources and distinctive boundary conditions other than continuity requirements. We are justified, therefore, in setting all h_{ik} to zero except for the more complicated h_{00}. Using conservation equation (3.5) and the z dependence of Δm during the forward phase of the oscillation, we find that h_{00} will grow if $z < z_1$. But if $z < z_1$ the equation is simpler since Δm is now zero. For $z < z_1$ the field change h_{00} can be shown to satisfy Eq. (2.3) just as f did for the unperturbed beam. But $g_{00} = f + h_{00}$ and it is convenient to choose $g_{00}(x=0) = 1$ since $g_{00} \equiv 1$ in a vacuum and this gives a uniformly scaled transformation of Minkowski space Eq. (I,2). However, $f(x=0) = 1$ and therefore $h_{00}(x=0) = 0$. As a solution to Eq. (2.3) h_{00} becomes identically zero for $z < z_1$. Hence, we conclude that all h_{ik} are zero for $z < z_1$. The essential point here is that if we look at the effect of the forward phase of the oscillation in the region $z < z_1$, we find that the h_{ik} field does not propagate into this region and the effect is nonexistent.

B. Rearward moving phase of oscillator

The nonzero four-velocity components of $\Delta m(t,x,z)$ moving opposite to the beam are $\hat{u}^0 = 2/f$ and $\hat{u}^3 = -1$. It should be noted that unless $x = 0$ this four-velocity does not obey the geodesic equations. However, if $z_2 - z_1$ is sufficiently small this four-velocity choice is a reasonable approximation. Using these $\hat{u}^i$ components, the nonzero components of $\hat{T}_{ik}$ are

$$\hat{T}_{00} = \Delta m f\,, \quad \hat{T}_{03} = 2\Delta m\,, \quad \hat{T}_{33} = 4\Delta m/f\,. \qquad (6.1)$$

Before discussing the inhomogeneous field equations generated by these terms, the homogeneous normal coordinate field equations (3.30)–(3.32) for H_{12}, H_{22}, and H_{23} can be quickly handled. Nonzero solutions for H_{12} and H_{22} were discussed in Sec. V and both can now be taken to be zero without a loss of generality. The conservation equation (3.7) removes all substantive variations of H_{23} inside of the beam. Extending it to all x, H_{23} may be taken to be zero. The three inhomogeneous equations (3.33)–(3.35) can be simplified by the conservation equation (3.8), where m is now independent of t. In the $z_1 < z < z_2$ region this yields

$$\Delta m(t,x,z) = \epsilon f(x) r(x) s[t - 2z/f(x)]\,, \qquad (6.2)$$

where ϵ is a first-order constant and r and s are zero-order arbitrary functions of their arguments. As was discussed with Eq. (5.5), the form of the argument of s corresponds to the null rearward velocity of the oscillating light. Our inhomogeneous normal coordinate equations in the $z_1 < z < z_2$ region become

$$\Box H_{33} = 16\epsilon r s\,, \qquad (6.3)$$

$$\Box H_{13} = -32\epsilon \int \frac{\partial rs}{\partial x} dz \tag{6.4}$$

with rs being zero outside of this region, and

$$\begin{aligned} \Box H_{03} &= 0\,, \quad z < z_1\,, \\ &= -4\epsilon f r[s(t,z) - s(t,z_1)]\,, \quad z_1 < z < z_2\,, \\ &= -4\epsilon f r[s(t,z_2) - s(t,z_1)]\,, \quad z_2 < z\,. \end{aligned} \tag{6.5}$$

Conservation equation (3.6) demands that H_{13} also satisfies

$$m \frac{\partial H_{13}}{\partial z} = \frac{2}{f} \frac{\partial f}{\partial x} \Delta m\,. \tag{6.6}$$

In analyzing the behavior of H_{13}, we note that for $z > z_2$ and for $|x| < X$ the right side of Eq. (6.6) is zero and this in Eq. (6.4) gives

$$\frac{\partial^2 H_{13}}{\partial x^2} = -32\epsilon \int_{z_1}^{z_2} \frac{\partial rs}{\partial x} dz\,, \quad z_2 < z\,. \tag{6.7}$$

Therefore H_{13} is essentially z independent in this region. In demanding continuity in x the z independence may be extended to the entire $z > z_2$ region. Hence H_{13} shares the time independence of its source much as f shared the time independence of m. In the region where $z < z_1$ and $|x| < X$, the right side of Eq. (6.4) is zero. By Eq. (6.6) we then find that $\partial^2 H_{13}/\partial x^2 = 0$ in this region. Demanding continuity in x, H_{13} can be no more than a function of t for $z < z_1$, which, to maintain causality, can be set to zero. Although H_{13} is more complicated for $z_1 < z < z_2$, the essential point is that due to an oscillation in the $z_1 < z < z_2$ region, H_{13} is consistent with zero for $z < z_1$, grows for $z_1 < z < z_2$, and then propagates unchanged with a beam cross section for $z > z_2$. This is similar to our conclusion at the end of Sec. VI A, in that, if we concentrate on the effect for $z < z_1$, we may ignore the contribution of some field components. In this case we may ignore H_{13}.

Normal coordinates H_{33} and H_{03} remain. Our task is to satisfy Eqs. (6.3) and (6.5) for some r and s. Simple conditions on H_{33} and H_{03} do not exist as they did for H_{23} and H_{13}. We again choose mX^2 to be small so that the approximation of Eq. (5.8) holds. The homogeneous solutions for $\Psi(t,x,z)$ given by Eq. (5.1) and the matched solutions $\psi_{\tilde{x},n}(x)$ [Eqs. (5.7) and (5.9)] can be used to fabricate a solution to the inhomogeneous equations.

One way to attack the two inhomogeneous equations is to first find the response function to a source of frequency ω, which has a set x displacement ξ and proceeds back along the entire z axis. Once obtained we then proceed to a delta function source in the temporal variable. The procedure involves expanding the inhomogeneous solution in terms of a complete set of homogeneous solutions $\psi_{\tilde{x},n}$ for the chosen ω. The requirement for a source at $x = \xi$ demands that $\tilde{x} = \xi$ and the use of the complete set $\{\psi_{x,n}\}$. An extended source over a range of x would require a comparable range of complete sets. We will therefore restrict our interest to those cases in which the x dependence can be approximated by a delta function. This restriction can be traced back to the existence of $f(x)$ in the d'Alembertian operator. Specifically, our equations are not fully separable. This effectively couples the variables t and z. This can be seen in the relationship of ω and κ in the homogeneous solutions. There ω and κ are independent, but, once chosen, their ratio determines the confinement distance $\tilde{x}$. In the inhomogeneous case the source position ξ must be identified with $\tilde{x}$ and this, in turn, determines ω/κ. Hence only ω or κ can be chosen freely. In this discussion we will fix ω and let the ratio determine κ. This means that $\omega t + \kappa z = \omega[t + 2z/f(\xi)]$ and t and z are coupled during the rearward directed phase of the oscillation in the single variable

$$\tau = t + [2/f(\xi)]\,z\,, \tag{6.8}$$

which is the natural temporal variable to use during the rearward phase. The response function is given by a Fourier transform and the resulting contour integration. Introducing $H(\tau,x)$ as a general inhomogeneous solution, our underlying equation to solve is

$$\Box H = \delta(x-\xi)\cos(\omega\tau + \varphi)U(\tau)\,, \tag{6.9}$$

where U is the unit step function with $U=0$ for $\tau<0$. Choosing a contour that ensures causality (which in this case means no response previous to the passage of the source), we find

$$\begin{aligned} H(\tau,x) = U(\tau) \sum_n \frac{\psi_{\xi,n}(\xi)\psi_{\xi,n}(x)}{4\omega_n(\omega_n^2 - \omega^2)M_n(\xi)} \\ \times\{\omega_n\cos(\omega\tau+\varphi) - \omega_n\cos(\omega_n\tau)\cos\varphi \\ + \omega\sin(\omega_n z)\sin\varphi\}\,, \end{aligned} \tag{6.10}$$

with M_n given by Eq. (5.6).

It should be noted that any source described by $\Delta m(\tau)$, independent of t, is one which moves rearward at the local speed of light at $x = \xi$. This source moves along the entire z axis and the solution H describes a wave which trails behind the source much as a flag will trail behind its pole. But, as mentioned before, the assumption that the light element remains at $x = \xi$ is an unrealistic one in that a null geodesic at ξ directed toward the rear would start to move toward the beam's axis. However the actual problem that we wish to consider is for light caught between two mirrors at z_1 and z_2. This problem does not have the above complaint because we may choose $z_2 - z_1$ small enough so that the inward tendency of the geodesics can be as small as desired and can be further compensated for by slight inclinations of the mirrors.

As discussed in Sec. VI A the forward phase of the oscillation produces no effect in the $z < z_1$ region. Even so, it would seem that the removal of this forward part of the oscillation would produce a gap in the temporal description of the rearward phase of the oscillation. We can see that this is not the case if we follow a cross section of Δm. As a cross section proceeds from z_2 to z_1 the quantity t increases uniformly. At z_1 the cross section is reflected. Since the retarded time t parametrizes forward-directed cross sections, the quantity t remains constant until reaching z_2. At z_2 the cross section is reflected back toward z_1 and t once again increases uniformly. Therefore, Δm is a continuous function of t for a fixed z during the rearward motion. In addition, this produces a source with a natural frequency inversely proportional to $z_2 - z_1$; namely,

$$\omega = \pi f(\xi)/(z_2 - z_1)\,. \tag{6.11}$$

It may be noted that the rearward directed Δm is indeed a step function of the variable τ during this oscillation.

Using a non-negative sinusoidal dependence for Δm, we choose

$$S(t,z) = s(\tau) = [1-\cos(\omega\tau)] \times [U(z-z_1) - U(z-z_2)] . \tag{6.12}$$

If we apply this to H_{33} we see from Eq. (6.3)

$$\Box H_{33} = 16\epsilon\delta(x-\xi)[1-\cos(\omega\tau)] \times [U(z-z_1) - U(z-z_2)] . \tag{6.13}$$

This may appear to be similar to Eq. (6.9) for H, but the source term no longer has the transient in τ and now has an explicit z dependence. The lack of a τ transient will be discussed shortly, but the z dependence is crucial in that it singles out specific positions along the z axis. It was the lack of a true z dependence, which allowed a solution for H by an expansion in terms of the homogeneous solutions. Each of those solutions represents a wave moving back along the entire beam axis passing all z values indiscriminately.

We are not, however, at an impasse. The wave guide effect provided by the light beam is beneficial. We recall that a sinusoidal source moving parallel to the z axis at $x=\xi$ and opposite to the beam's flux produces a wave that acts as an infinite tail to the source with constant intensity [Eqs. (6.9) and (6.10)]. If we were to stop or reflect the source at $z=z_1$ it is reasonable to assume that the gravitational wave would continue to propagate into the $z<z_1$ region. There will be some new transients in the wave front, but the structure of the infinite tail will not significantly change since the light beam acts as a wave guide for propagation into the $z<z_1$ region and there is no other dissipation in the system. Hence the wave of the semi-infinite track source should look like that of the wave of the infinite track source some distance behind the wave front.

If we now consider a sinusoidal oscillatory source between z_1 and z_2, we even lose the semi-infinite track source case. But, an observer between z_1 and z_2 will see a source that locally is indistinguishable from a sinusoidal infinite track source. Therefore, an observer at z in the $z<z_1$ region will see the gravitational wave of the preceding paragraph much as before if we also demand that $z \ll z_2$. The last condition ensures that the wave structure will approach that of the semi-infinite track case. In addition, by using a constant sinusoidal oscillating source, the transients in the semi-infinite track case can be ignored. In conclusion, we may use the steady-state wave of the infinite track case [Eqs. (6.9) and (6.10)] for our sinusoidal oscillator in the $z<z_1$ region if $z \ll z_2$. Hence, the solution to Eq. (6.13) for $z \ll z_1$ is

$$H_{33} = 4\epsilon \sum_{n=0} \frac{\psi_{\xi,n}(\xi)\psi_{\xi,n}(x)}{\omega_n^2 M_n(\xi)} \times \left[1 + \frac{\omega^2\cos(\omega_n\tau) - \omega_n^2\cos(\omega\tau)}{\omega_n^2-\omega^2}\right]. \tag{6.14}$$

The solution for H_{03} is more difficult because of the unique source of Eq. (6.5). But, by the discussion of t preceding Eq. (6.11), we find $s(t,z_1) = s(t,z_2)$. Hence the source term of Eq. (6.5) becomes zero for $z>z_2$. By introducing a new function

$$F = \frac{\partial H_{03}}{\partial z}, \tag{6.15}$$

the resulting equation is manageable. That is, F satisfies

$$\Box F = -8\epsilon\omega\sin(\omega\tau)\delta(x-\xi)[U(z-z_1) - U(z-z_2)] .$$

Using Eqs. (6.9) and (6.10) we find

$$F = 2\epsilon\omega \sum_{n=0} \frac{\psi_{\xi,n}(\xi)\psi_{\xi,n}(x)}{M_n(\xi)(\omega_n^2-\omega^2)} \times [(\omega/\omega_n)\sin(\omega_n\tau) - \sin\omega\tau] . \tag{6.16}$$

We may also show that H_{33} is related to H_{03} through

$$F = -\frac{1}{4} f(\xi)\frac{\partial H_{33}}{\partial z}. \tag{6.17}$$

C. Antenna structure and response

The effect of the above oscillatory source solution on the beam itself is similar to that of the homogeneous solution discussed in Sec. V B. The oscillation in energy density exists, but it is minute. However, there is the possibility of amplification with an antenna. Considering the general principal that the best antenna for a given radiator has the same basic structure as the radiator itself, we can create an antenna out of an oscillating light element. The mirrors can be placed at z_3 and z_4 with $z_3 < z_4 \ll z_1$. The simplest case is when $z_4 - z_3 = z_2 - z_1$. The antenna also shares the radiator's $x=\xi$ position and, hence, has the same natural frequency.

The phase relation between a light element in the antenna and a gravitational wave is important. We start the cycle at z_4 and follow the light element. Since the gravitational wave was formed by a source at $x=\xi$ it propagates back at the same null speed as the antenna light element at $x=\xi$. Hence, during this rearward propagation the light element and the wave are phase locked and the interaction is constant during this passage. This is somewhat analogous to the situation of a surfer riding a wave. If the surfer stays in phase with the wave the wave–surfer interaction is also constant. But a difference exists in that the surfer is a dissipative system, while the gravitational wave case is not. The time integral of the wave–antenna light interaction may grow. At z_3 the light element reflects forward and the effect of the wave on it becomes identical to that of the wave on a beam element as it progresses from z_3 to z_4. This was discussed previously and the small resulting oscillation in energy density is negligible. What is important is the phase relation between the light and wave upon being reflected at z_4. Let us measure the retarded times in units of $2(z_4 - z_3)/f(\xi)$. This is the time interval in t units to go from z_4 to z_3 (or in τ units to go from z_3 to z_4). If the light element starts at z_4 at $(t,\tau) = (0,0)$, then the values at z_3 will be $(t,\tau) = (1,0)$. Upon completion of the first cycle at z_4, $(t,\tau) = (1,1)$. This is identical to the light element in the radiator and, hence, to the gravitational wave itself since $z_4 - z_3 = z_2 - z_1$. Therefore, the wave and light element start the second cycle in the antenna with the same phase relation as they had initially. Again the interaction is maintained during the rearward propagation to z_3 and the effect will continue to build. After N cycles any effect will be N times that of a single cycle.

Consider the photon energy in the antenna. The energy change in the antenna is given by the eikonal equation. If the four-velocity of an observer is v^i then the local optical frequency of a light element with four-velocity u^i is

$$\nu = \nu_0 g_{ik} u^i v^k , \tag{6.18}$$

where ν_0 is a constant for a given light element in the antenna. The photon energy is obtained by multiplying by Planck's constant.

During our cycle u^i, v^i, and g_{ik} all change. But we may take the observer to be a mirror of the antenna. Then the change in the mirror's velocity δv^i due to the wave passage is periodic with the same period as that of the light element in the antenna. Hence δv^i is zero after one antenna cycle. The same holds for h_{ik} as observed at z_3 or z_4. Therefore only δu^i has a net change and to first order we have $u^0 = 2/f + \delta u^0$ and $u^3 = -1 + \delta u^3$. Our eikonal relation after one period becomes

$$\delta \nu = \nu_0 f^{-1/2} (f \delta u^0 + \delta u^3) .$$

The fractional change in frequency after N periods is

$$\delta \nu / \nu = N f^{-1/2} \mathscr{F} , \tag{6.19}$$

where

$$\mathscr{F}(\tau) = f \delta u^0 + \delta u^3$$

and τ selects the specific light element in the first period of oscillation. As discussed, the change δu^i in one period is equal to the change during the rearward passage; i.e.,

$$\delta u^i = \int_{z_4}^{z_3} \frac{du^i}{ds} ds ,$$

with τ constant. Changing variables, we have to first order

$$\delta u^i = \int_{z_3}^{z_4} \frac{du^i}{ds} dz , \quad \tau \text{ constant} .$$

By the surfing analogy the effect of the interaction du^i/ds is independent of z in the antenna for any light element. Therefore,

$$\delta u^i = \frac{du^i}{ds} (z_4 - z_3) .$$

This can be evaluated by the geodesic equations and written in terms of the h_{ik}. These in Eq. (6.19) give

$$\mathscr{F}(z) = -\left[\frac{\partial H_{03}}{\partial z} + \frac{f}{2} \frac{\partial H_{33}}{\partial z} + \frac{2}{f^2} \left(\frac{\partial h_{00}}{\partial t} - f \frac{\partial h_{00}}{\partial z} \right) - \frac{2}{f^2} \frac{df}{dx} h_{01} \right] (z_4 - z_3) .$$

By Eqs. (6.8), (6.15), and (6.17) this simplifies to

$$\mathscr{F}(z) = \left[F + \frac{2}{f^2} \left(\frac{\partial h_{00}}{\partial \tau} + \frac{df}{dx} h_{01} \right) \right] (z_4 - z_3) . \tag{6.20}$$

As long as $\mathscr{F}$ is nonzero, the frequency change given by Eq. (6.19) will grow linearly with time and the light elements in this optical antenna will undergo linearly increasing red or blue shifts.

Further computation of $\mathscr{F}$ is tedious. But since h_{00} and h_{01} can be derived from H_{03} and H_{33}, they too can be written in terms of F and its derivatives. The specifics of this system are not as important here as the approximate magnitude of the red/blue shifts. So we will approximate Eq. (6.20) by

$$\mathscr{F} \approx F(z_4 - z_3) , \tag{6.21}$$

and consider applications.

An astrophysical application seems unlikely due to the necessity of sufficient identity between the radiator and the antenna. As discussed in Sec. V B, beam powers of a few hundred terawatts have been reported for a pulsed x-ray laser. If a somewhat continuous beam could be constructed with a power P of 10^{14} W with $X = 1$ m then by the discussion in Sec. V B the fundamental frequency $\omega_0 \approx 10^{27}$ Hz. With our radiator the actual gravitational frequency is given by Eq. (6.11), where $z_2 - z_1$ (at the smallest) might be the thickness of a film 1 Å thick. Hence $\omega \approx 10^{19}$ Hz and our system is highly subresonant.

To find an order of magnitude for the effect of such a gravitational wave, we see by Eq. (6.16) that only the first term in the expansion for F would be of any significance. Hence,

$$|F| \approx (\epsilon \omega / \omega_0^2) [\psi_{\xi,0}^2 / M_0(\xi)] \, c \Delta x , \tag{6.22}$$

where c had been previously set to unity and Δx is the x thickness of the radiator around $x = \xi$. By Eqs. (5.10) and (6.2),

$$\frac{\epsilon}{m} \approx \frac{P_r}{P} \frac{X}{\Delta x} , \tag{6.23}$$

where P_r is the power of the light in the radiator per $2X$ in the y direction. Equations (5.6), (5.11), (6.22), and (6.23) give

$$|F| \approx 2.5 \times 10^{-60} \omega P_r .$$

Using $|F|$, Eqs. (6.11), (6.19), and (6.21), and also noting that $z_4 - z_3 = z_2 - z_1$, we find that

$$\delta \nu / \nu \approx 10^{-50} N P_r ,$$

where P_r is in watts [Eq. (6.23)] and N is the number of oscillations in the antenna. The relation holds for any highly subresonant system and is independent of the antenna power, ω, X, and P (as long as $P \gg P_r$). The time T for N oscillations is $T = 2\pi N / \omega$ and the time to yield a fractional change $\delta \nu / \nu$ is

$$T \approx (\delta \nu / \nu) \, P_r^{-1} (z_2 - z_1) \, 10^{42} \text{ sec} , \tag{6.24}$$

where P_r is in watts and $z_2 - z_1$ is in meters. Say that we could create a radiator with an optical power $P_r \approx 10^{-2} P = 10^{12}$ W with $z_2 - z_1 \approx 1$ Å as discussed above. If we could also detect a shift in the optical frequency in the antenna of $\delta \nu / \nu = 10^{-10}$, then by Eq. (6.24) $T \approx 10^{10}$ sec or a few hundred years. The rapid development of the x-ray laser may reduce this value quickly. The recent report[12] of an increase in the power by six orders of magnitude would, for example, reduce the resonant frequency ω_0 by three orders of magnitude. We would, therefore, still have a highly subresonant system, and the above analysis holds. With the parameters used here and keeping $P_r \approx 10^{-2} P$, Eq. (6.24) states that T diminishes by the six orders of magnitude to $T \approx 10^4$ sec. Considering, among other things, that the beam of light should be continuous for at least this length of time at a power P and that the thin film radiator must hold together with a power P_r inside of it, a laboratory attempt at detection would appear to be prohibitive.

Although true detection of this type of gravitational wave appears to be unlikely, the system is rather unique and may allow selective application. For example, it is not unreasonable to imagine a cosmological situation in the early universe where an immense directed energy transport exists of the kind envisioned here. The channeling of gravitational waves back along the beam axis may provide a useful coupling of regions.

[1]H. A. Lorentz, *Problems of Modern Physics* (Dover, New York, 1967), p. 210.
[2]R. C. Tolman, P. Ehrenfest, and B. Podolsky, Phys. Rev. **37**, 602 (1931).
[3]R. C. Tolman, *Relativity, Thermodynamics and Cosmology* (Clarendon, Oxford, 1934), pp. 272–288.
[4]M. O. Scully, Phys. Rev. D **19**, 3582 (1979).
[5]W. B. Bonnor, Commun. Math. Phys. **13**, 163 (1969).
[6]R. W. Nackoney, J. Math. Phys. **14**, 1239 (1973).
[7]R. W. Nackoney, J. Math. Phys. **18**, 2146 (1977).
[8]A. Banerjee, J. Math. Phys. **16**, 1188 (1975).
[9]K. S. Thorne, Rev. Mod. Phys. **52**, 285 (1980).
[10]R. W. Nackoney "Gravitational wave solutions of a perturbed light beam" (unpublished).
[11]A. L. Robinson, Science **215**, 488 (1982).
[12]Sci. Am. **253** (1), 58 (1985).

Three remarks on Powers' theorem about irreducible fields fulfilling CAR

Klaus Baumann
Department of Physics, Princeton University, Princeton, New Jersey 08544

(Received 7 April 1986; accepted for publication 30 April 1986)

First it is shown that within a relativistic Fermi field theory, a bound $\|\Psi_k(f,t)\| \leqslant C \|f\|_2$ already implies canonical anticommutation relations (CAR). Then under Powers' assumptions a linear, first-order differential equation for the fields $\psi_k(x,t)$ is derived. This shows that in the set of generalized free fields fulfilling CAR only the free fields are irreducible at time zero. Finally Fermi fields in two space-time dimensions are considered. It is shown that only four-fermion interaction might be compatible with CAR and a bound on the coupling strength is derived.

I. INTRODUCTION

In 1967 Powers[1] showed that a relativistic field theory for fermions fulfilling canonical anticommutation relations (CAR) and in which the time-zero fields form an irreducible set can never describe any interaction if the dimension of space-time is at least three.

We want to supplement his result by the following remarks.

(1) If instead of CAR we assume only that the fields at a fixed time are bounded operators and their norm $\|\psi_k(f,t)\|$ is bounded by the L_2-norm $\|f\|_2$ of the test functions, then CAR follows from locality (see Sec. II).

(2) We show that the Fermi fields are not only generalized free, but that these fields fulfill a linear, first-order, partial differential equation. This shows also that in the set of generalized free fields fulfilling CAR only the free fields are irreducible (see Sec. III).

(3) Finally we consider two-dimensional Fermi field theories. First we derive that the only interaction that might be compatible with CAR is the four-fermion interaction proposed by Thirring,[2] but, in addition, we get a bound on the coupling strength. As a consequence, we show that a spin-$\frac{1}{2}$ Majorana field is a free field and that there cannot exist any nontrivial solution for the massless Thirring model that fulfills CAR. To our knowledge none of the known solutions for nontrivial, two-dimensional fermionic models does fulfill CAR (see Ref. 3). Unfortunately we are not able to disprove the existence of such solutions. For the Yukawa interaction there exist solutions that fulfill CAR (and CCR), but this model contains fermions and bosons and therefore does not fit into our scheme (see Sec. IV).

Let us recall the assumptions that led to Powers' ICAR theorem.

(i) Relativistic quantum field theory: ψ is a m-component Fermi field in the axiomatic frame given by Wightman. At spacelike distances the fields anticommute, i.e.,

$$\{\psi_k(x),\psi_l(y)\} = 0 = \{\psi_k(x)^*,\psi_l(y),\} \quad \text{if } (y-x)^2 < 0.$$

The number of space-time dimensions is $n+1$.

(ii) Canonical anticommutation relations (CAR): (a) For fixed time t the fields $\psi_k(f,t)$ and their adjoints $\psi_k(f,t)^*$ define operators if smeared with test functions $f \in \mathscr{S}(\mathbf{R}^n)$.

(b) The fields fulfill CAR:

$$\{\psi_k(f,t),\psi_l(g,t)\} = 0 = \{\psi_k(f,t)^*,\psi_l(g,t)^*\},$$

$$\{\psi_k(f,t)^*,\psi_l(g,t)\} = \delta_{kl}\int_{\mathbf{R}^n}(f^*g)(x)\,d^nx.$$

(iii) Irreducibility: The smeared fields at a fixed-time act irreducibly on the Hilbert space $\mathscr{H}$.

(iv) Existence of $\partial_t\psi$: We assume that for $f,g \in \mathscr{S}(\mathbf{R}^n)$ the expressions

$$\text{(a)}\quad \lim_{t\to 0}(1/t)[\psi(f,t)^\# - \psi(f,0)^\#]\Omega = (\partial_t\psi)(f,0)^\#\Omega,$$

$$\text{(b)}\quad \lim_{t\to 0}(1/t)[\psi(f,t)^\# - \psi(f,0)^\#]\psi(g,0)^\#\Omega = (\partial_t\psi)(f,0)^\#\psi(g,0)^\#\Omega$$

converge strongly.

Throughout this paper $\psi(f,t)^\#$ stands for either $\psi(f,t)$ or $\psi(f,t)^*$ and we often drop the indices labeling the field components if they are of no special importance. If we do not write a time argument we always mean the field operator at time zero.

II. CONSEQUENCES OF A L_2-BOUND FOR $\|\psi(f,t)\|$

It is well known that from

$$\{\psi_k(f,t)^*,\psi_k(f,t)\} = \int_{\mathbf{R}^n}|f|^2(x)\,d^nx, \quad k=1,\dots,m,$$

the bound $\|\psi_k(f,t)\| = \|f\|_2 = \|\psi_k(f,t)^*\|$ follows. Now we shall show that such a bound

$$\|\psi_k(f,t)^\#\| \leqslant C\|f\|_2, \quad k=1,\dots,m, \tag{2.1}$$

already implies that the equal time commutation relations are determined by the two-point functions and their singularities are at most δ functions.

Theorem 1: The assumptions (i) relativistic quantum field theory (rel. QFT) and (iii) (irreducibility) supplemented by the bound (2.1) imply

$$\{\psi_k(x,t)^\#,\psi_l(y,t)\} = (\Omega,\{\psi_k(x,t)^\#,\psi_l(y,t)\}\Omega) = C_{\psi_k^\#\psi_l}\delta(y-x). \tag{2.2}$$

Proof: (a) Because Lemma 1 in Powers' paper[1] is the key to this whole business we write down his result as the following lemma.

Lemma 2.1: For $f,g,h \in L_2(\mathbb{R}^n)$, we have

$$\|[\psi(h)^\#\{\psi(g)^\#,\psi(f,t)\}]\| \leqslant C_n\|h\|_2 \sup|f| \sup|g|\,|t|^n, \tag{2.3}$$

where n is the number of space dimensions and C_n is a constant.

(b) For $t = 0$ we get from the above lemma

$$[\psi(h)^\#\{\psi(g)^\#,\psi(f)\}] \equiv 0, \quad \text{for } f,g \in \mathscr{S}(\mathbb{R}^n) \quad \text{and} \quad h \in L_2(\mathbb{R}^n).$$

Now $\{\psi(g)^\#,\psi(f)\}$ is a bounded operator by (2.1) and therefore irreducibility (iii) implies

$$\{\psi_k(g,0)^\#,\psi_l(f,0)\} = (\Omega,\{\psi_k(g,0)^\#,\psi_l(f,0)\}\Omega).$$

Also because time translation is given by a unitary operator e^{iHt}, which leaves the vacuum Ω invariant, this is true for all t.

(c) *Lemma 2.2:*

$$\{\psi_k(x,t)^\#,\psi_l(y,t)\}(y-x)_j \equiv 0, \quad \text{for } j = 1,\dots,n.$$

Proof: Take $f,g \in \mathscr{D}(\mathbb{R}^n)$ contained in a cube of side length L. To simplify notation we drop the indices k and l. Like Powers[1] we define for $\underset{\sim}{k} \in \mathbb{Z}^n$ and $t > 0$ the Hermitian projections $E^t_{\underset{\sim}{k}}$ by

$$(E^t_{\underset{\sim}{k}} f)(x) = \begin{cases} f(x), & \text{if } x_i \in [k_i t,(k_i+1)t\,], \quad i = 1,\dots,n, \\ 0, & \text{otherwise.} \end{cases} \tag{2.4}$$

With their help we get by linearity

$$\left\|\iint \{\psi(x,0)^\#,\psi(y,t)\}(y-x)_j\, g^\#(x) f(y)\, d^n x\, d^n y\right\| = \left\|\sum_{\underset{\sim}{k},\underset{\sim}{l} \in \mathbb{Z}^n} \iint \{\psi(x,0)^\#,\psi(y,t)\}(y-x)_j\, g^\#(x) f(y) E^t_{\underset{\sim}{k}}(x) E^t_{\underset{\sim}{l}}(y)\, d^n x\, d^n y\right\|. \tag{2.5}$$

This sum is finite because f and g have compact support! By locality we get $|k_i - l_i| \leqslant 1$, for all $i = 1,\dots,n$. If we write

$$y_j - x_j = (y_j - k_j t) + (k_j t - x_j) \tag{2.6}$$

and use the triangle inequality, we have

$$\leqslant \sum_{\substack{\underset{\sim}{k},\underset{\sim}{l} \\ |\underset{\sim}{k}-\underset{\sim}{l}| \leqslant 1}} \left\|\iint \{\psi(x,0)^\#,\psi(y,t)\}[E^t_{\underset{\sim}{k}}(x) g^\#(x) E^t_{\underset{\sim}{l}}(y)(y_j - k_j t) f(y) + E^t_{\underset{\sim}{k}}(x)(k_j t - x_j) g^\#(x) E^t_{\underset{\sim}{l}}(y) f(y)]\, d^n x\, d^n y\right\|. \tag{2.7a}$$

Now we can use the bound (2.1) and get

$$\leqslant 4 \times 3^{n+1}(L+2)^n \max|f| \max|g|\,|t|. \tag{2.7b}$$

The additional factor $|t|$ is a consequence of the factor $(y-x)_j$ in (2.5) and is due to the fact that $|y_j - k_j t|$ and $|k_j t - x_j|$ can never exceed $3|t|$!

As t goes to zero Lemma 2.2 follows from (2.5) and (2.7). This shows that the singularity in the equal-time anticommutator is at most of δ function and therefore the second part of the theorem has been proved.

Remark 2.3: In Theorem 1 we do not end up with the usual CAR algebra, because we did not use Lorentz covariance. A possible way to incorporate Lorentz covariance is to use a Källen–Lehmann representation[4] for the two-point function. As an example let us take a charged spin-$\frac{1}{2}$ field, which we assume to have a definite behavior under P (space reflection) and C (charge conjugation). The Källen–Lehmann representation has the form

$$(\Omega,\{\bar\psi(x,t),\psi(y,s)\}\Omega) = \int_0^\infty dm^2\{\rho_1(m^2)(-i)S(y-x,s-t;m) + \rho_2(m^2) i\Delta(y-x,s-t;m^2)\}, \tag{2.8}$$

with $\bar\psi = \psi^* \gamma_0$ and $\rho_1(m^2) \geqslant 0$, $2m\rho_1(m^2) \geqslant \rho_2(m^2) \geqslant 0$. (We stick to Schweber's notation[5] for ease of reference.) The bound (2.1) implies

$$\int_0^\infty dm^2 \rho_1(m^2) \leqslant C^2. \tag{2.9}$$

After rescaling we get the usual CAR algebra.

III. A FIRST-ORDER DIFFERENTIAL EQUATION FOR $\psi(x,t)$

Let us outline the idea: If $(\partial_t \psi)(f,t)$ exists then because of locality the support of $\{\psi_k(x,t)^*,(\partial_t \psi_l)(y,t)\}$ consists only of the point $x = y$ and the bound $\|\psi(f,t)\| = \|f\|_2$ implies that this anticommutator is a linear combination of $\delta(y-x)$ and

$\nabla_j\,\delta(y-x), j=1,...,n$, but by CAR we write these terms as $\{\psi_k(x,t)^*,\psi_k(y,t)\}$ and $\{\psi_k(x,t)^*,\nabla_j\,\psi_k(y,t)\}$. From irreducibility we then conclude that $\partial_t\psi_l$ equals a linear combination of $\psi_k^\#$ and $\nabla_j\,\psi_k^\#$.

Because many of these steps are already included in Powers' paper,[1] let us recapitulate his ICAR theorem: For $n\geqslant 2$ we have the following equations:

$$(\partial_t\,\psi_k)(f,t)=\sum_{l=1}^{m}\psi_l(T_{kl}f,t)+\psi_l(\hat{T}_{kl}f,t)^*,\quad k=1,...,m, \tag{3.1}$$

where T_{kl} are linear and $\hat{T}_{kl}$ antilinear operators from $\mathscr{S}(\mathbf{R}^n)$ into $L_2(\mathbf{R}^n)$.

The followng theorem characterizes T_{kl} and $\hat{T}_{kl}$ as first-order differential operators with constant coefficients.

Theorem 2: Under the assumption of the ICAR theorem we have, for all $k,l=1,...,m$ and $f\in\mathscr{S}(\mathbf{R}^n)$,

$$T_{kl}f=(M_{kl}+\boldsymbol{\alpha}_{kl}\nabla)f, \tag{3.2a}$$

$$\hat{T}_{kl}f=(N_{kl}+\boldsymbol{\beta}_{kl}\nabla)f^*, \tag{3.2b}$$

where M_{kl}, N_{kl}, $\boldsymbol{\alpha}_{kl}$, and $\boldsymbol{\beta}_{kl}$ are constants.

Proof: (a) From CAR it is obvious that T_{kl} is uniquely determined by $\{\psi_l(g)^*,(\partial_t\psi_k)(f)\}$ and $\hat{T}_{kl}$ by $\{\psi_l(g),(\partial_t\psi_k)(f)\}$.

(b) The following lemma shows that $\{\psi_l(x,0)^\#,(\partial_t\psi_k)(y,0)\}$ is a linear combination of $\delta(y-x)$ and $\nabla_j\,\delta(y-x)$, $j=1,...,n$.

Lemma 3.1:

$$\{\psi_l(x,0)^\#,(\partial_t\,\psi_k)(y,0)\}(y-x)_i(y-x)_j\equiv 0,\quad \text{for all } i,j=1,...,n. \tag{3.3}$$

Proof: For $f,g\in\mathscr{D}(\mathbf{R}^n)$ consider

$$\iint\{\psi(x,0)^\#,\psi(y,t)\}(y-x)_j\,g^\#(x)f(y)d^nx\,d^ny.$$

By using the same steps as explained in the proof of Lemma 2.2, we get the esimate

$$\left\|\iint\{\psi(x,0)^\#,\psi(y,t)\}(y-x)_i(y-x)_j\,g^\#(x)f(y)d^nx\,d^ny\right\|\leqslant 8\cdot 3^{n+2}(L+2)^n\max|g|\max|f|\,|t|^2. \tag{3.4}$$

Dividing by t and passing to the limit $t=0$ we get Lemma 3.1.

(c) Therefore we get from CAR

$$\begin{aligned}\{\psi_l(x,0)^*,(\partial_t\,\psi_k)\,(y,0)\}&=M_{kl}\delta(y-x)-\boldsymbol{\alpha}_{kl}\nabla\delta(y-x)\\&=M_{kl}\{\psi_l(x)^*,\psi_l(y)\}-\boldsymbol{\alpha}_{kl}\{\psi_l(x)^*,(\nabla\psi_l)\,(y)\}.\end{aligned} \tag{3.5}$$

This proves (3.2a) and (3.2b) follow along the same line and from the fact that $\psi_l(\hat{T}_{kl}f,0)^*$ has to be linear in f.

Remark 3.2: We expect this linear differential equation for $\psi(x,t)$ to reduce to a free-field equation, if we take Lorentz covariance and eventually certain discrete symmetries into account. As already pointed out in Remark 1.3, a Källen–Lehmann representation is a feasible method to prove this. We want to illustrate this for the most interesting case, namely the example of a charged spin-$\frac{1}{2}$ field.

We start with the Källen–Lehmann representation (2.8), CAR requires $\int_0^\infty dm^2\,\rho_1(m^2)=1$ and the existence of $\partial_t\,\psi_k(f,t)\Omega$ implies $\int_0^\infty dm^2\,\rho_1(m^2)m^2<\infty$.

Consider now the equal-time anticommutator

$$\begin{aligned}\{\bar{\psi}(x,t),(i\gamma^\mu\,\partial_\mu\psi)\,(y,t)\}&=\int_0^\infty dm^2\,\rho_1(m^2)\,(-i)\left[\underbrace{(i\gamma^\mu\,\partial_\mu-m)S(y-x,0;m)}_{\equiv 0}+mS(y-x,0;m)\right]\\&\quad+\int_0^\infty dm^2\,\rho_2(m^2)i(i\gamma^\mu\,\partial_\mu)\Delta(y-x,0;m)\\&=\underbrace{\int_0^\infty dm^2\,\rho_1(m^2)m}_{M_1}\,\gamma_0\delta(y-x)+\underbrace{\int_0^\infty dm^2\,\rho_2(m^2)}_{M_2}\,\gamma_0\delta(y-x).\end{aligned} \tag{3.6}$$

Therefore we have the equation

$$\{\bar{\psi}(x,t),(i\gamma^\mu\partial_\mu-(M_1+M_2))\psi(y,t)\}=0. \tag{3.7}$$

From irreducibility (iii) we derive the free-field equation

$$(i\gamma^\mu\partial_\mu - (M_1+M_2))\psi(y,t) = 0. \tag{3.8}$$

Therefore the spectral weight function ρ_1 is of the form $\rho_1(m^2) = C_1\delta(m^2 - (M_1+M_2)^2)$ and CAR implies $C_1 = 1$. By definition, $M_1 = \int_0^\infty dm^2\,\rho_1(m^2)m = M_1 + M_2$ and therefore $M_2 = 0$, which implies $\rho_2 \equiv 0$ because ρ_2 is a positive distribution.

As a by-product we have shown the following corollary.

Corollary 3.3: Within the set of generalized free fields given by the Källen–Lehmann representation (2.8), fulfilling CAR and the bound

$$\int_0^\infty dm^2\, m^2\rho_1(m^2) < \infty, \tag{3.9}$$

only the free fields of a definite mass M fulfill the irreducibility assumption (iii).

IV. WHAT HAPPENS IN ONE SPACE DIMENSION?

This case is of special interest because, as already mentioned in the Introduction, there are some explicitly solvable models available (see Ref. 3). None of the known solutions fulfills CAR. We are not able to show that only free fields can fulfill CAR, but, nevertheless, we hope our analysis gives some new insight into this problem.

In the case $n = 1$ we cannot conclude from Lemma 2.1 that $[\psi(h)^\#\{\psi(g)^\#,(\partial_t\,\psi)(f)\}]$ vanishes. The situation improves if we consider higher commutators, e.g., for the fourfold commutator we get the following lemma.

Lemma 4.1: There exists a constant $\hat{C}$ such that for $h\in L_2(\mathbb{R})$ and $f_1,\dots,f_4\in\mathscr{D}(\mathbb{R})$ we have

$$\|[\psi(h)^\#\{\psi(f_4)^\#[\psi(f_3)^\#\{\psi(f_2)^\#,\psi(f_1,t)^\#\}]\}]\| \leqslant \hat{C}\,\|h\|_2 \prod_{i=1}^{4} \max|f_i|\,|t|^2. \tag{4.1}$$

Proof: Goes along the same line as Lemma 1 in Powers ICAR theorem.

The detailed structure of the operator $\{\psi(f_4)^\#[\psi(f_3)^\#\{\psi(f_2)^\#,(\partial_t\,\psi)(f_1)\}]\}$ will be displayed by the following lemma.

Lemma 4.2: For $f_1,\dots,f_4\in\mathscr{D}(\mathbb{R})$ we have

(i) $$\{\psi_{k_4}(f_4)^\#[\psi_{k_3}(f_3)^\#\{\psi_{k_2}(f_2)^\#,(\partial_t\,\psi_{k_1})(f_1)^\#\}]\} = (\Omega,\{\psi_{k_4}(f_4)^\#[\psi_{k_3}(f_3)^\#\{\psi_{k_2}(f_2)^\#,(\partial_t\,\psi_{k_1})(f_1)^\#\}]\}\Omega)$$
$$= C(\psi^\#_{k_4},\psi^\#_{k_3},\psi^\#_{k_2},\psi^\#_{k_1})\int_{\mathbb{R}}(f_4^\# f_3^\# f_2^\# f_1^\#)(x)\,dx, \tag{4.2}$$

where $C(\psi^\#_{k_4},\dots)$ is a complex number, whose absolute value is bounded by $C_0 \leqslant 216$.

(ii) $C(\psi^\#_{k_4},\psi^\#_{k_3},\psi^\#_{k_2},\psi^\#_{k_1})$ is totally antisymmetric under permutations of the fields $\psi^\#_{k_i}$ and it vanishes if two arguments equal each other.

Proof: (a) From Lemma 2.1 for the case $n = 1$ we get the boundedness of the triple commutation (i) and it equals its vacuum expectation value because of Lemma 4.1 and irreducibility (iii).

(b) Using the same methods as in proof of Lemma 2.2 one can easily obtain for $f_1,\dots,f_4\in\mathscr{D}(\mathbb{R})$ the estimate

$$\left\|\int \{\psi(x_4)^\#[\psi(x_3)^\#\{\psi(x_2)^\#,\psi(x_1,t)^\#\}]\}(x_j - x_i)f_4(x_4)^\# f_3(x_3)^\# f_2(x_2)^\# f_1(x_1)^\#\,dx_4\cdots dx_1\right\| \leqslant \check{C}\prod_i \max|f_i|\,|t|^2 \tag{4.3}$$

and therefore we have

$$\{\psi(x_4)^\#[\psi(x_3)^\#\{\psi(x_2)^\#,(\partial_t\,\psi)(x_1)^\#\}]\}(x_j - x_i) \equiv 0. \tag{4.4}$$

This together with locality and translation invariance proves the representation (4.2).

(c) If we take the characteristic function E of the unit interval as test function for the fields, then

$$|C(\psi^\#_{k_4}\cdots)| = \left\|\lim_{t\to 0}\frac{1}{t}\{\psi_{k_4}(E)^\#)[\psi_{k_3}(E^t_{l_3})^\#\{\psi_{k_2}(E^t_{l_2})^\#,\psi_{k_1}(E^t_{l_1},t)^\#\}]\}\right\|$$
$$< \lim_{t\to 0}\frac{1}{t}\left\|\sum_{\substack{l_1\cdots l_4\\ |l_2-l_1|<1\\ |l_3-l_1|<1\\ |l_4-l_1|<1}}\{\psi_{k_4}(E^t_{l_4})^\#[\psi_{k_3}(E^t_{l_3})^\#\{\psi_{k_2}(E^t_{l_2})^\#,\psi_{k_1}(E^t_{l_1},t)^\#\}]\}\right\|$$
$$< \lim_{t\to 0}(1/t)2^3\cdot 3^3[(1/t)+2]t^2 = 216. \tag{4.5}$$

A more refined estimate yields immediately $C_0 < 80$. This proves part (i) of the lemma.

(d) If we apply the Jacobi identities

$$[A\{B,C\}] + [B\{C,A\}] + [C\{A,B\}] = 0 \tag{4.6}$$

and

$$\{A[B,C]\} + \{B[A,C]\} + [C\{A,B\}] = 0 \tag{4.7}$$

to the triple commutator (4.2) and use CAR we get

$$C(\psi^{\#}_{k_4},\psi^{\#}_{k_3},\psi^{\#}_{k_2},\psi^{\#}_{k_1}) = (-1)C(\psi^{\#}_{k_4},\psi^{\#}_{k_2},\psi^{\#}_{k_3},\psi^{\#}_{k_1}) = (-1)C(\psi^{\#}_{k_3},\psi^{\#}_{k_4},\psi^{\#}_{k_2},\psi^{\#}_{k_1}). \tag{4.8}$$

The antisymmetry under exchange of $\psi^{\#}_{k_2}$ and $\psi^{\#}_{k_1}$ follows from

$$0 = \partial_t(\psi_{k_2}(f_2,t)^{\#},\psi_{k_1}(f_1,t)^{\#}\} = \{(\partial_t\,\psi_{k_2})\,(f_2,t)^{\#},\psi_{k_1}(f_1,t)^{\#}\} + \{\psi_{k_2}(f_2,t)^{\#},(\partial_t\,\psi_{k_1})\,(f_1,t)^{\#}\}. \tag{4.9}$$

(e) For real test functions the triple commutator (4.2) is symmetric under permutation of the test functions and antisymmetric under permutation of the field labels. Therefore $C(\psi^{\#}_{k_4},...)$ has to vanish if two field labels are equal.

This completes the proof of the lemma.

The next lemma shows that with the help of these coefficients $C(\psi^{\#}_{k_4},...)$ we can uniquely express the double commutator $[\psi(h)^{\#}\{\psi(g)^{\#},\partial_t\,\psi(f)^{\#}\}]$ as a linear combination of all the fields ψ_l and ψ^{*}_l.

Lemma 4.3: For $f_1,f_2,f_3\in\mathscr{D}(\mathbb{R})$ we have

$$[\psi_{k_3}(f_3)^{\#}\{\psi_{k_2}(f_2)^{\#},(\partial_t\,\psi_{k_1})(f_1)^{\#}\}] = \sum_{l=1}^{m} C(\psi^{*}_l,\psi^{\#}_{k_3},\psi^{\#}_{k_2},\psi^{\#}_{k_1})\psi_l(f^{\#}_3 f^{\#}_2 f^{\#}_1) + \sum_{l=1}^{m} C(\psi_l,\psi^{\#}_{k_3},\psi^{\#}_{k_2},\psi^{\#}_{k_1})\psi_l((f^{\#}_3 f^{\#}_2 f^{\#}_1)^{*})^{*}. \tag{4.10}$$

Proof: (a) Let Q be the difference of the right- and left-hand side of Eq. (4.10). For $g\in\mathscr{D}(\mathbb{R})$ it follows from Lemma 4.1 and from CAR that, for all $l=1,...,m$,

$$\{\psi_l(g),Q\} = 0 = \{\psi_l(g)^{*},Q\}. \tag{4.11}$$

By continuity this equation can be extended to all $g\in L_2(\mathbb{R})$.

(b) Because all odd n-point functions vanish there exists a unitary and Hermitian operator U_I such that

$$U_I\,\Omega = \Omega,\quad U_I\,\psi(f,t)U_I^{-1} = -\psi(f,t). \tag{4.12}$$

From (4.12) we get immediately

$$[\psi_l(g)^{\#},U_IQ\,] = -\,U_I\{\psi_l(g)^{\#},Q\} = 0. \tag{4.13}$$

Therefore we conclude from irreducibility (iii)

$$U_IQ = (\Omega,U_IQ\Omega) = (\Omega,Q\Omega) = 0 \tag{4.14}$$

because Q is an odd monomial. But U_I is unitary and therefore $Q\equiv 0$.

Before we draw conclusions let us summarize what we already know about $\{\psi(g)^{\#},(\partial_t\,\psi)(f)\}$. From Lemma 3.1 we get

$$\{\psi_l(x,t)^{\#},(\partial_t\,\psi_k)(y,t)\} = \hat{C}_{\psi^{\#}_l,\psi_k}(x,t)\delta(y-x) + B_{\psi^{\#}_l,\psi_k}(x,t)\delta'(y-x). \tag{4.15}$$

We claim that the matrix B is constant, because estimate (2.5) proves the boundedness of

$$\left|\left|\iint\{\psi(x)^{\#},(\partial_t\,\psi)\,(y)\}(y-x)g^{\#}(x)f(y)dx\,dy\right|\right|$$

and it is easy to show that for $h\in L_2(\mathbb{R})$

$$\left[\psi(h)^{\#},\iint\{\psi(x)^{\#},(\partial_t\,\psi)\,(y)\}\times(y-x)g^{\#}(x)f(y)dx\,dy\right]\equiv 0. \tag{4.16}$$

By extracting the vacuum expectation value, we can rewrite (4.15) as

$$\{\psi_l(x,t)^{\#},(\partial_t\,\psi_k)\,(y,t)\} = A_{\psi^{\#}_l,\psi_k}\delta(y-x) + B_{\psi^{\#}_l,\psi_k}\delta'(y-x) + C_{\psi^{\#}_l,\psi_k}(x,t)\delta(y-x), \tag{4.17}$$

where the matrices A and B are uniquely determined by the vacuum expectation value of (4.17).

Corollary 4.4: Let us consider a spin-$\frac{1}{2}$ Majorana field

$$\psi_1(f,t)^{*} = \psi_1(f^{*},t),\psi_2(f,t)^{*} = \psi_2(f^{*},t)$$

and

$$\{\psi_k(f,t),\psi_l(g,t)\} = \frac{1}{2}\,\delta_{kl}\int_{\mathbb{R}}(fg)\,(x)\,dx. \tag{4.18}$$

There are only two different fields and therefore $C(\psi_{k_4},...,\psi_{k_1})\equiv 0$ which in turn implies

$$\{\psi_k(f),(\partial_t\,\psi_l)\,(g)\} = (\Omega,\{\psi_k(f),(\partial_t\,\psi_l)(g)\}\Omega). \tag{4.19}$$

Proceeding in the same way as Powers[1] did, we end up with the free-field equations

$$\partial_t\,\psi_1 + \partial_x\,\psi_1 + m\psi_2 = 0,$$
$$\partial_t\,\psi_2 - \partial_x\,\psi_2 - m\psi_1 = 0. \tag{4.20}$$

This confirms the old statement by Thirring,[2] that a spin-$\frac{1}{2}$ Majorana field can never describe any interaction in a canonical theory.

Remark 4.5: For a charged spin-$\frac{1}{2}$ field (or, equivalently, for two spin-$\frac{1}{2}$ Majorana fields), there is up to permutations only one nontrivial triple commutator

$$\{\psi_2(f_4)^{*}[\psi_2(f_3)\{\psi_1(f_2)^{*},(\partial_t\,\psi_1)(f_1)\}]\} = C(\psi^{*}_2,\psi_2,\psi^{*}_1,\psi_1)\int_{\mathbb{R}}(f^{*}_4 f_3 f^{*}_2 f_1)(x)dx. \tag{4.21}$$

Let us define $\lambda = iC(\psi^{*}_2,\psi_2,\psi^{*}_1,\psi_1)$; then λ is real and Lemma 4.2 provides the bound $|\lambda|\leqslant C_0\leqslant 216$. How can we interpret (4.21)? In terms of Wick polynomials we could reproduce (4.21) by identifying

$$i\,\partial_t\psi_1 \equiv \lambda\!:\psi^{*}_2\,\psi_2\,\psi_1\!: + \text{ linear terms in } \psi, \tag{4.22}$$

and this would also explain formula (4.10). Unfortunately we do not know *a priori* how to define such operator products. Otherwise we could use the arguments by Powers[1] and the ideas of Sec. III to derive the Thirring equation

$$\begin{aligned} i(\partial_t + \partial_x)\psi_1 + m\psi_2 &= \lambda\!:\psi^{*}_2\,\psi_2\,\psi_1\!:,\\ i(\partial_t - \partial_x)\psi_2 + m\psi_1 &= \lambda\!:\psi^{*}_1\,\psi_1\,\psi_2\!:, \end{aligned} \tag{4.23}$$

supplemented by CAR and a bound on λ, which is very strange. Another peculiarity is described by the following corollary.

Corollary 4.6: $m=0$ implies $\lambda=0$, i.e., in our framework there exists no solution to the massless Thirring model as long as $\lambda \neq 0$, which fulfills CAR.

Proof: For $m = 0$ we get for the anticommutator

$$\{\psi_2(x,t)^*, i(\partial_t + \partial_x)\psi_1(y,t)\}$$
$$= \lambda\{\psi_2(x,t)^*, :\psi_2^* \psi_2 \psi_1:(y,t)\}. \quad (4.24)$$

By construction [remember the explanations leading to Eq. (4.17)!] the vacuum expectation value of the right-hand side in Eq. (4.24) vanishes. [A nonzero expectation value $c\delta(y-x)$ would produce a mass term $c\{\psi_2(x,t)^*, \psi_2(y,t)\}$ because of CAR!] Therefore we get the equations

$$(\Omega, \{\psi_k(x,t)^\#, i(\partial_t + \partial_x)\psi_1(y,t)\}\Omega) = 0,$$
$$(\Omega, \{\psi_k(x,t)^\#, i(\partial_t - \partial_x)\psi_2(y,t)\}\Omega) = 0, \quad (4.25)$$

and from a Källen–Lehmann representation [see Eqs. (3.6) and (3.7)] we see immediately that $\rho_1(m^2) = \delta(m^2)$ and $\rho_2(m^2) = 0$. From this we conclude that ψ is a free-spinor field and λ has to vanish.

Remark 4.7: Finally let us consider a theory with more than two spin-$\frac{1}{2}$ Majorana multiplets, e.g., a Gross–Neveu model with at least three Majorana spinors or a Federbush model, which is described in terms of two charged spin-$\frac{1}{2}$ fields.

Lemma 4.2 tells us that only a four-fermion interaction might be compatible with CAR. There is the possibility of having different "coupling constants" C_i, because there are more than four different fields available, but all these C_i have to fulfill an *a priori* bound $|C_i| < C_0 < 216$.

For the Federbush model, which is explicitly solvable, all known solutions do not fulfill CAR.

ACKNOWLEDGMENT

It is a great pleasure to thank A. S. Wightman for many discussions on this subject.

The author was supported by a grant from the Max Kade Foundation, New York.

[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] W. E. Thirring, "A soluble relativistic field theory," Ann. Phys. (NY) **3**, 91 (1958).

[3] S. N. M. Ruijsenaars, "On the two-point functions of some integrable relativistic quantum field theories," J. Math. Phys. **24**, 922 (1983).

[4] I. Raszillier and D. H. Schiller, "Spectral representations for any spin," Nuovo Cimento A **48**, 617 (1967).

[5] S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory*, (Harper and Row, New York, 1962), pp. 671–675.

Four-dimensional boson field theory. II. Existence

George A. Baker, Jr.
Theoretical Division, Los Alamos National Laboratory, University of California, Los Alamos, New Mexico 87545

(Received 15 April 1985; accepted for publication 30 April 1986)

The existence of the continuum, quantum field theory found by Baker and Johnson [G. A. Baker, Jr. and J. D. Johnson, J. Phys. A **18**, L261 (1985)] to be nontrivial is proved rigorously. It is proved to satisfy all usual requirements of such a field theory, except rotational invariance. Currently known information is consistent with rotational invariance however. Most of the usual properties of other known Euclidean boson quantum field theories hold here, in a somewhat weakened form. Summability of the sufficiently strongly ultraviolet cutoff bare coupling constant perturbation series is proved as well as a nonzero radius of convergence for high-temperature expansions of the corresponding continuous-spin Ising model. The description of the theory by these two series methods is shown to be equivalent. The field theory is probably not asymptotically free.

I. INTRODUCTION AND SUMMARY

Recently Baker and Johnson[1] reported a solution to a long-standing problem. The problem was the construction of a nontrivial, scalar, self-interacting boson quantum field theory in four dimensions. In the context of usual approaches the existence of such a theory was virtually excluded. Attention has been centered on the $g_0{:}\phi^4{:}_4$ theory because higher polynomial powers in the interaction are not renormalizable.[2] The usual tool in quantum field theory is the cutoff or regularized perturbation theory and the approach has been to adjust the bare parameters of the theory or add counterterms in such a way as to leave the resultant physical quantities predicted by the theory finite. Of course, as is pointed out by careful treatises[2] on the subject, it is conceivable that one can construct other theories that are nonrenormalizable, but these require an infinite number of counterterms that in principle require an infinite number of experimental measurements to fix such a theory. From the theoretical point of view this approach is not too pleasant because of the complexity of the specifications required to yield even a finite physical theory. A key theoretical method to overcome this latter difficulty was foreshadowed by Simon[3] who realized that for $P(\phi)_2$ field theory, the rate of falloff of the (vacuum subtracted) two-point Schwinger function at large distances dominated the higher-order (vacuum subtracted) Schwinger functions. Thus the control of the two-point function for such a theory suffices to control the entire theory and leads to a finite physical theory. Nevertheless Newman[4] proved, under mild assumptions, that such a field theory cannot exist if its degree (which corresponds formally to the degree of the polynomial interaction in the Hamiltonian) exceeds 4, which would appear to leave the search for a nontrivial self-interacting boson theory in four dimensions with only $g_0{:}\phi^4{:}_4$ as a candidate. However recent numerical[5] and theoretical[6,7] evidence strongly suggests that $g_0{:}\phi^4{:}_4$ theory is itself trivial, i.e., a generalized free-field that has no scattering. This situation is rather mysterious because, for the corresponding classical case (nonlinear Klein–Gordon equation) nontrivial scattering is known[8] to occur and by the correspondence principle one might expect this feature not to be disrupted by quantization.

Although in recent years, the main thrust of quantum field theory has been along the lines of non-Abelian gauge theories, the standard model requires the Higgs boson (four real fields in the simplest case) to transform massless gauge fields into massive ones. Hence the construction of a four-dimensional, nontrivial, self-interacting boson quantum field in the Higgs sector is certainly still a necessary component of the theory.

A fresh way to attack the solution of this construction problem was discovered by Baker[9] in the study of limitations of critical index universality. Baker and Johnson[1] call this the "method of phantom fields." It bears a relationship to the spirit of local effective Lagrangian theory[10] and the theory of ultraviolet renormalons.[11–13] In the latter theory it was found[14,15] that they are proportional to the insertion of local irrelevant variables, as, for example, ϕ^6, ϕ^8, etc. The method of phantom fields contemplates polynomial interactions of arbitrary degree and so the whole family has an infinite number of parameters. In this way it accords with the abstract theory of nonlinear wave equations in being able to deal mathematically with a wide and rich class of model theories and leaves to the physics of the problem the choice of the relevant ones.

In this paper I demonstrate that for a given interaction polynomial and a given four-dimensional Euclidean, space-lattice ultraviolet cutoff that there exists a limiting process that defines a continuum random field that *exists* for any prescribed positive physical mass m. This field satisfies the axioms of Nelson's reconstruction theorem.[16] An exception is that rotational invariance is not proved, but only shown to be consistent with currently available information. By Nelson's theorem, a Minkowski space theory can then be constructed that satisfies the Wightman axioms[17] of quantum field theory. The other results reported by Baker and Johnson[1] concerning the *nontriviality* and, in principle, *computability* by series methods of the continuum field theory depend on numerical estimation procedures. These results will be discussed separately.[18] The theories so constructed are probably not asymptotically free. I draw the reader's attention to the point that the computability question for the class of continuum phantom field theories contains in it the notorious, currently unsolved by rigorous methods, question of the construction of $g_0{:}\phi^4{:}_4$ field theory. The simpler question of the rigorous computability of a sufficiently strongly cutoff lattice field theory is answered here in the affirmative.

The methods employed in this paper are to associate the lattice cutoff Euclidean quantum field theory with a corresponding continuous-spin Ising model. Then the machinery of rigorous statistical mechanics, consisting mainly of inequalities and convergent series expansions, is used to establish the existence and various properties of the field theories in both the infinite volume and continuum limits.

In the second section, I describe the phantom field model and make the connection with the corresponding continuous-spin Ising model. In the course of this connection a coefficient K is introduced to multiply the site–site coupling term. This parameter plays the role of the inverse temperature in statistical mechanics, and expansions in this parameter, or quantities closely related to it, are often termed high-temperature expansions. The Nelson[16] axioms needed in his reconstruction theorem are discussed and it is shown that, except for rotational invariance, they are all valid for models of this structure.

In the third section, I assemble relevant information about the infinite volume or thermodynamic limit at fixed lattice spacing, a. These results are either simply cited with appropriate references or a proof is given or sketched as is required. All the results in this section assume that the lattice spacing is positive and not zero. A few of them require the additional hypothesis that the physical mass m, which describes the exponential decay at large distances, is positive as well. Three boundary conditions are treated: free boundary conditions, periodic boundary conditions, and Dirichlet boundary conditions. The thermodynamic limit of the free energy per unit volume exists for all these boundary conditions and must agree independently of the boundary conditions. The thermodynamic limit for the Schwinger functions (the expectation value of a product of field monomials at various lattice sites) with Dirichlet boundary conditions exists. The doubly cutoff (finite volume and nonzero lattice spacing) perturbation series in powers of the coefficient λ_0 of the interaction polynomial is summable to the correct physical answer for all positive real λ_0 and defines correctly the infinite volume limit by an appropriate limiting process. The free energy is uniformly continuous in λ_0 for $0\leqslant\lambda_0<\infty$. The physical mass is defined as the rate of exponential decay of the unsubtracted two-point Schwinger function in the limit of infinite separation of the points, and a pseudomass is introduced for finite volume that is proved to be equal to the physical one in the thermodynamic limit. Nelson's reflection positivity property holds for these models where the reflections are through lattice symmetry planes and his result that the corresponding transfer matrix can be expressed as the exponential of a nonpositive, Hermitian operator holds as well. Various fundamental correlation function inequalities hold such as the Girffiths inequality, the Griffiths–Kelly–Sherman inequality, the Fortuin–Kasteleyn–Ginibre inequality, the reflection positivity inequality, and the Frölich–Lieb inequality. There are many consequences. The two-point Schwinger function is monotonically decreasing with the separation of the two points and is also log convex at least in directions perpendicular to planes of lattice reflection symmetry. The Sokal cluster property holds. This result is a key one because it shows that if the two-point Schwinger function decays exponentially with distance, then so too, modulo a few powers of the distance, do the higher-order Schwinger functions. Thus, if the two-point Schwinger function is controlled, then by two-point dominance the whole theory is also controlled.

In addition if we suppose that the physical mass $m>0$, as well as that the lattice spacing $a>0$, then we can show that the Schwinger functions are uniformly continuous in λ_0 and are independent of the boundary conditions. In addition a variant definition of the mass (second moment definition) and the amplitude renormalization constant Z_3 are also uniformly continuous in λ_0. To show that these results are not vacuous, I use the continuous-spin Ising model formulation and show that there exists a $\widehat{K}$ such that for $K<\widehat{K}$ (sufficiently high temperature) the physical mass is positive and goes to infinity as $K\to 0$. In addition I show that for K sufficiently small that the high-temperature series converges and so for λ_0 non-negative real the free energy and the Schwinger functions are analytic in some disk centered at $K=0$. To close the third section I show that if the interaction polynomial is of degree $2p$ then there exists a K_0 such that if for $|K|<K_0$ and $0<|\widetilde{\lambda}_0|<\lambda_m$, $|\arg\widetilde{\lambda}_0|\leqslant(p+1)\pi/2$, then the free energy per unit volume is analytic in $\widetilde{\lambda}_0$.

In the fourth section, I discuss the continuum limit. By a Peierls-type argument these models must display long-range order for sufficiently low temperature (large K). By my definition, the occurrence of long-range order forces the physical mass to zero. For a finite box the pseudomass is continuous in K. I can therefore select a sequence of boxes with lattice spacings that go to zero simultaneously with the box size going to infinity and chose a K for each box in the sequence to keep the pseudomass fixed. By amplitude renormalization, this process defines a continuum limit that yields a mass and amplitude renormalized field theory. By two-point dominance, uniformly in the sequence of box sizes, the whole set of Schwinger functions are also controlled. Thus I have shown that theories of this type do exist, and satisfy all of Nelson's axioms, save rotational invariance. I discuss this last point further in the last section.

In the fifth and final section I discuss the perturbation series in the coefficient $\widetilde{\lambda}_0$ of the interaction polynomial. I show that for K sufficiently small (high enough temperature), that is, for a sufficiently strong, lattice ultraviolet cutoff, the asymptotic series in $\widetilde{\lambda}_0$ is generalized Borel summable to the physically correct answer. By analytic continuation this summation implies, at least in the neighborhood of $K=0$, the physically correct answer. A consequence of this result is that the full, lattice cutoff $\widetilde{\lambda}_0$ series implies the full K-series for any positive real $\widetilde{\lambda}_0$. Contrariwise, as we mentioned in our summary of Sec. IV as the K-series is convergent for any $\widetilde{\lambda}_0$ and small enough K. A knowledge of this expansion suffices to determine the complete $\widetilde{\lambda}_0$

expansion, for K small enough. However again analytic continuation (here in K) gives us the complete results and so mathematically the two series expansion methods contain equivalent descriptions of this family of field theories.

As an illustration, I compute the first-order, lattice cutoff perturbation expansion terms for the four-line g and the six-line λ renormalized coupling constants for the case where the interaction polynomial is of degree 6. This example illustrates that a phantom field can be constructed that has, to leading order, $g<0$, where $g \geqslant 0$ is proven for $g_0{:}\phi^4{:}_4$ theory. The naive continuum limit $a \to 0$ of this theory has an apparent $g_0<0$ that would look unstable; however, the cutoff version has a nonzero coefficient of the ϕ^6 term to maintain stability.

By considering the interchange of limits so that instead of summing the $\widetilde{\lambda}_0$ series first and then taking the continuum limit, which leads to the physically correct model, one can look at the coefficients of the $\widetilde{\lambda}_0$ expansion term by term, and consider their continuum limit. Several features are then evident. First the coefficients are, of course, not finite term by term even after mass and amplitude renormalization. Second it is easy to exhibit terms in the two-particle scattering amplitude that do not go to zero as the particle momentum becomes large, so the theory is probably not asymptotically free. Third there are terms which depend on the lattice cutoff and differ for different lattices so the resultant theory is probably cutoff dependent. Finally, even though the continuum limit of the series terms is not necessarily finite, they are rotationally invariant in momentum space, so the theory probably is rotationally invariant and so permits the construction of a family of boson quantum field theories that satisfy the Wightman axioms.

II. EUCLIDEAN, POLYNOMIAL, BOSON QUANTUM FIELD THEORY MODEL

The models I wish to study are closely related to the scaling limit of the continuous-spin Ising model. A great many of their properties have been reviewed by Baker.[19] I will work in four-dimensional Euclidean space because, if we are successful in satisfying Nelson's axioms[16] then by his reconstruction theorem a Minkowski space-field theory satisfying the Wightman axioms[17] can be obtained. To guide the imagination we start with a structure,

$$Z=\int \mathscr{D}\phi(x)\exp\left\{-\int d^4x\left[(\nabla\phi)^2 + m_0^2\phi^2+\lambda_0 \not{h}(\phi)\right]\right\}. \tag{2.1}$$

Because of the behavior of the parameters in the model with the ultraviolet cutoff, (2.1) can be viewed as a symbolic shorthand for a set of limiting procedures that I will subsequently detail.

The first step in my approach is to introduce a finite portion of an infinite space lattice, with a field variable on each lattice site. This procedure introduces both a volume and an ultraviolet cutoff. I thus start from

$$Z=m^{-1}\int_{-\infty}^{+\infty}\cdots\int\prod_{\mathbf{r}} d\phi_{\mathbf{r}} \times\exp\left\{-\sum_{\mathbf{r}} v\left[\frac{8}{q}\sum_{\{\boldsymbol{\delta}\}}\frac{(\phi_{\mathbf{r}}-\phi_{\mathbf{r}+\boldsymbol{\delta}})^2}{a^2} + m_0^2\phi_{\mathbf{r}}^2+\lambda_0{:}\not{h}(\phi_{\mathbf{r}}){:}\right]\right\}, \tag{2.2}$$

where M is a formal normalization constant, $\mathbf{r}$ ranges over a finite portion of the space lattice, $\{\boldsymbol{\delta}\}$ is one-half the set of nearest-neighbor sites on the lattice, v is the specific volume per lattice site, e.g., a^4 for the hyper-simple-cubic lattice, a is the lattice spacing, q is the lattice coordinate number, $\not{h}(\phi_{\mathbf{r}})$ is a lower, semibounded, even, monic polynomial of degree $2p$, and ${:}\phi^n{:}$ is the normal-ordered product. As long as the lattice spacing is greater than zero, the normal-ordered product[19] is

$${:}\phi^{2p}{:}=\sum_{j=0}^{p}\frac{(2p)!(-1)^j}{(2p-2j)!j!}2^{-j}C^j\phi^{2p-2j}, \tag{2.3}$$

where C is the Fock-space commutator $[\phi^-,\phi^+]$ that, for the hyper-simple-cubic lattice in a box of edge L with N^4 sites, is

$$C=\frac{1}{(2\pi)^4}\left(\frac{2\pi}{L}\right)^4\sum_{\mathbf{k}}\left[m_0^2+4a^{-2}\sum_{\{\boldsymbol{\delta}\}}\sin^2\left(\frac{1}{2}\mathbf{k}\cdot\boldsymbol{\delta}\right)\right]^{-1}, \tag{2.4}$$

where $\mathbf{k}$ ranges in steps of $(2\pi/L)$ over a cube $-\pi/a<k_i<\pi/a$, $i=1,\ldots,4$. In the limit as the box size becomes infinite, C goes over into an integral over $\mathbf{k}$. This limit having been taken, if we now let $a\to 0$ then $C\to\infty$ in a manner proportional to a^{-2}. By the use of (2.3) I can reexpress (2.2) as

$$Z=M^{-1}\int_{-\infty}^{+\infty}\cdots\int\prod_{\mathbf{r}} d\sigma_{\mathbf{r}}\exp\left\{K\sum_{\mathbf{r}}\sum_{\{\boldsymbol{\delta}\}}\sigma_{\mathbf{r}}\sigma_{\mathbf{r}+\boldsymbol{\delta}} -\sum_{\mathbf{r}}\left[\widetilde{A}\sigma_{\mathbf{r}}^2+\widetilde{\lambda}_0 P(\sigma_{\mathbf{r}})\right]\right\}, \tag{2.5}$$

where M is a different formal normalization constant and P is again a lower semibounded, even, monic polynomial of degree $2p$, provided $a>0$. The notation of (2.5) is given by

$$\sigma_{\mathbf{r}}=\phi_{\mathbf{r}}(16v/qa^2K)^{1/2},\quad \widetilde{A}=\tfrac{1}{2}K(1+a^2m_0^2),$$
$$\widetilde{\lambda}_0=\lambda_0(qKa^2/16v)^p v,\quad P(x)=\sum_{j=1}^{p}a_jx^{2j},\quad a_p=1. \tag{2.6}$$

The introduction of the extra parameter K in (2.5) allows us to impose the following normalization on the scale of $\sigma_{\mathbf{r}}$:

$$1=\langle 1\rangle=\langle\sigma_{\mathbf{r}}^2\rangle|_{K=0} =\frac{\int_{-\infty}^{+\infty}x^2\exp\{-\widetilde{A}x^2-\widetilde{\lambda}_0P(x)\}dx}{\int_{-\infty}^{+\infty}\exp\{-\widetilde{A}x^2-\widetilde{\lambda}_0P(x)\}dx}, \tag{2.7}$$

which determines $\widetilde{A}$ as a function of $\widetilde{\lambda}_0$ and P. We will defer discussion of the boundary conditions to a latter section.

I now introduce the concept of "phantom fields," which is central to my approach. I choose $\widetilde{\lambda}_0=O(1)$ and $a_j=O(1)$ with respect to the lattice spacing a. By means of the relation between $\sigma_{\mathbf{r}}$ and $\phi_{\mathbf{r}}$ of (2.6) and (2.3) and (2.4) we may readily compute that the coefficient of ϕ^{2n} in (2.2) will be proportional to $a^{2(n-2)}$. In other words the coefficient of ϕ^4 is of order unity, that of ϕ^6 is of order a^2, ϕ^8 of order a^4, and so on. This means that if we were to write out the continuum limit, symbolic shorthand without further attention, no polynomial term beyond ϕ^4 would seem to appear. Nevertheless, it is far too hasty to jump to the conclusion that the presence of σ^6, say, in (2.5) does not affect the continuum limit. We call these fields "phantom fields" because they appear to vanish in the continuum limit. Even so, that they may affect the continuum limit can be seen by the following illustration. The normal-ordered products from the phantom fields contribute coefficients of order unity to the coefficient of ϕ^4. For example,

$$a^2{:}\phi^6{:} = a^2\phi^6 - 15a^2C\phi^4 + 45a^2C^2\phi^2 - 15a^2C^3.$$

Since $C \propto a^{-2}$ we get a negative contribution to the coefficient of ϕ^4 in the continuum limit. With a judicious choice of $\not{h}$, I can produce a model, which is stable for large ϕ, i.e., the integral in (2.2) over $\phi_{\mathbf{r}}$ converges for $a > 0$, which has a positive (divergent) sign of the coefficient of ϕ^4 and no term visible to compensate it. This model avoids Newman's[4] proof of nonexistence because in his sense, the "polynomial degree" remains 4 for all phantom field models.

We close this section with a discussion of Nelson's[16] axioms for a continuum Euclidean field theory as they apply to lattice cutoff field theory. The axioms are: let there be given a random field $\phi(\mathbf{x})$ with the following properties [where $f^{(n)}(r_1,...,r_n)$ is a testing function, i.e., it is infinitely differentiable and vanishes outside a finite region of "space-time"].

(a) Euclidean invariance.

$$T_\eta \phi[f^{(n)}(\mathbf{r}_1,...,\mathbf{r}_n)] = \phi[f^{(n)}(O\mathbf{r}_1 + \mathbf{b},...,O\mathbf{r}_n + \mathbf{b})], \tag{2.8}$$

where O is an orthogonal transformation including reflections, and T_η is a member of a group of measure preserving transformations that are indexed by η, $\eta\mathbf{r} = O^{-1}(\mathbf{r} - \mathbf{b})$. This property holds for the lattice cutoff models in the restricted sense that it holds for the symmetry group of the lattice. If I impose periodic boundary conditions, then we get at once the translational invariance by any of the four fundamental lattice vectors. When the lattice spacing goes to zero, this property becomes translational invariance. As far as rotations and reflections go, only the lattice symmetries are built in and it remains to be shown that the full orthogonal group develops in the continuum limit.

(b) Markov property. This property is easily explained on the lattice. Let X be a finite subset of the points on the lattice and ∂X be all the nearest-neighbor points of the points in X that are not themselves points in X. Then, if u is any random variable that depends only on $\phi(\mathbf{x})$, $\mathbf{x}\in X$, the Markov property says

$$E\Big(u\Big|\textstyle\sum(X')\Big) = E\Big(u\Big|\textstyle\sum(\partial X)\Big), \tag{2.9}$$

where X' is the complement of X on the lattice and $\Sigma(X)$ is the set of all $\phi(\mathbf{x})$, where $\mathbf{x}\in X$. In words (2.9) says that the expected value of a random variable with support in X is not changed, once the values of the nearest-neighbor fields are fixed, by fixing further spins in the complement of X. This property of my model can be seen by inspection of (2.2) [or (2.5)] because once the $\phi(\mathbf{x})$ in the set ∂X are fixed, the expectation value integral factors into an integral over X and one over $X'/\partial X$. The second one cancels between the numerator and the denominator of the expectation value leaving only integrals over $\phi(\mathbf{x})$, $\mathbf{x}\in X$, irrespective of what spins are or are not fixed in $X'\setminus\partial X$, which means the Markov property is built into this model.

(c) Hermiticity.

$$\phi\{[f^{(n)}(\mathbf{r}_1,...,\mathbf{r}_n)]^*\} = \{\phi[f^{(n)}(\mathbf{r}_1,...,\mathbf{r}_n)]\}^*. \tag{2.10}$$

This property follows easily here because the measure is a non-negative definite function of real $\phi(\mathbf{x})$. For further details, the reader is referred to Baker[19] and Nelson.[16]

III. INFINITE VOLUME LIMIT

Much of the basic material needed in this section is already available for the models under consideration in Baker's review article.[19] I emphasize immediately that throughout this section a finite lattice spacing will be assumed and the passage to the continuum limit will be considered in the next section. The first step is to establish the existence of this limit and this step has been taken in the aforementioned reference for three sets of boundary conditions. Namely, first free boundaries (a subscript plus is used to denote them) where any term of the form $(\phi_{\mathbf{r}} - \phi_{\mathbf{r}+\boldsymbol{\delta}})^2$ in (2.2) that involves a ϕ which lies outside the finite portion of the lattice considered is dropped. Second, periodic boundary conditions (no subscript is used to denote them), where the portion of the lattice considered is infinitely repeated in a periodic manner so when a $\phi_{\mathbf{r}+\boldsymbol{\delta}}$ occurs that falls outside the portion considered, it is replaced by the corresponding ϕ on the opposite boundary. Finally Dirichlet boundaries conditions are used. (Here a subscript minus is used to denote them.) Here if a $\phi_{\mathbf{r}+\boldsymbol{\delta}}$ lies outside the portion considered, it is just taken to be zero. The main tools used are various correlation function inequalities. Next define the Schwinger functions,

$$S(\mathbf{r}_1,...,\mathbf{r}_n) = (MZ)^{-1}\int_{-\infty}^{+\infty}\!\!\cdots\!\int\Big[\prod_{\mathbf{r}} d\phi_{\mathbf{r}}\Big]\phi_{\mathbf{r}_1}\cdots\phi_{\mathbf{r}_n} \times\exp\bigg\{-\sum_{\mathbf{r}} v\bigg[\sum_{\{\boldsymbol{\delta}\}}\frac{(\phi_{\mathbf{r}} - \phi_{\mathbf{r}+\boldsymbol{\delta}})^2}{a^2} + m_0^2\,\phi_{\mathbf{r}}^2 + \lambda_0{:}\not{h}(\phi_{\mathbf{r}}){:}\bigg]\bigg\}, \tag{3.1}$$

where M, Z are as in (2.2). The appropriate boundary conditions, by use of (3.1), are implied to define not only S but S_+ and S_- as well. The conclusions are, in brief,

$$\begin{aligned} &Z_+ \geqslant Z \geqslant Z_- \geqslant 0, \\ &S(\mathbf{r}_1,...,\mathbf{r}_n) \geqslant S_-(\mathbf{r}_1,...,\mathbf{r}_n) \geqslant 0, \\ &S_+(\mathbf{r}_1,...,\mathbf{r}_n) \geqslant S_-(\mathbf{r}_1,...,\mathbf{r}_n) \geqslant 0, \end{aligned} \tag{3.2}$$

and independent of system size,

$$S(\mathbf{r}_1,...,\mathbf{r}_n) \leqslant \frac{\int_0^\infty x^n \exp[-m_0^2a^4x^2 - \lambda_0 a^4{:}\not{h}(x){:}]\,dx}{\int_0^\infty \exp[-(m_0^2a^4 + 8a^2)x^2 - \lambda_0 a^4{:}\not{h}(x){:}]\,dx}. \tag{3.3}$$

As to the variation with system size Z_+ decreases monotonically, since if two chunks of lattice A and B are joined

$$Z_{+,A+B} \leqslant Z_{+,A} Z_{+,B} \tag{3.4a}$$

and Z_- increases monotonically in the same sense, i.e.,

$$Z_{-,A+B} \geqslant Z_{-,A} Z_{-,B} \,. \tag{3.4b}$$

These inequalities permit the conclusions, if one treats a portion of the lattice of edge L

$$\frac{1}{L^4} \ln Z_+ \geqslant \lim_{L\to\infty} \frac{1}{L^4} \ln Z_+ \geqslant \lim_{L\to\infty} \frac{1}{L^4} \ln Z_- \geqslant \frac{1}{L^4} \ln Z_- \,. \tag{3.5}$$

Equation (3.5) gives meaning to the "free energy" per unit volume as

$$f_\pm = -\lim_{L\to\infty} (1/L^4) \ln Z_\pm \tag{3.6}$$

has been proved to exist for fixed lattice spacing for every non-negative, bare-couple constant λ_0 and bare mass m_0^2. Further it has been proved[19] that

$$f = -\lim_{L\to\infty} (1/L^4) \ln Z = f_- \,, \tag{3.7}$$

and

$$\mathscr{S}(\mathbf{r}_1,...,\mathbf{r}_n) = \lim_{L\to\infty} \inf S(\mathbf{r}_1,...,\mathbf{r}_m) \,,$$

$$\geqslant \mathscr{S}_-(\mathbf{r}_1,...,\mathbf{r}_n) = \lim_{L\to\infty} S_-(\mathbf{r}_1,...,\mathbf{r}_n) \,. \tag{3.8}$$

These results establish the full existence of the infinite volume limit for Dirichlet boundary conditions and at least partial results for other boundary conditions.

The next step is to see what can be concluded from the cutoff perturbation series in λ_0. Here, of course, I keep the lattice spacing $a > 0$ and I will first start with a finite sized box of edge L. First notice that since $:\mathscr{P}(\phi):$ is lower semibounded, there exists a constant Y such that $:\mathscr{P}(\phi): + Y \geqslant 0$, for all real ϕ. If I replace $:\mathscr{P}:$ by $:\mathscr{P}: + Y$ in (2.2) then Z is replaced by $Z \exp(-\lambda_0 v N^4 Y)$, where there are N^4 lattice sites in our box of edge L. It is now immediate that

$$Z \exp(-\lambda_0 v N^4 Y) = \int_0^\infty e^{-\lambda_0 s} \, d\rho(s) = g(\lambda_0) \,, \tag{3.9}$$

where $d\rho$ is a non-negative measure. Functions of this structure can be exploited by means of the following generalized Padé approximants[20,21]:

$$B_{n,-1}(z) = \sum_{j=1}^{n} \alpha_j \exp(-z\sigma_j) \,,$$
$$B_{n,0}(z) = \hat{\alpha}_0 + \sum_{j=1}^{n} \hat{\alpha}_j \exp(-z\hat{\sigma}_j) \,, \tag{3.10}$$

where the α_n, $\hat{\alpha}_j$, σ_j, $\hat{\sigma}_j$ are never negative and are determined by the equations

$$g(z) - B_{n,-1}(z) = O(z^{2n}) \,,$$
$$g(z) - B_{n,0}(z) = O(z^{2n+1}) \,. \tag{3.11}$$

These approximants have the properties

$$B_{n,-1}(z) \leqslant B_{n+1,-1}(z) \leqslant g(z) \leqslant B_{n+1,0}(0) \leqslant B_{n,0}(z) \,, \tag{3.12}$$

for real, non-negative z. For every such z, (3.12) implies that as $n\to\infty$ both $B_{n,-1}(z)$ and $B_{n,0}(z)$ converge to a limit (not necessarily the same limit). Using these approximants and the monotonicity properties in box size established in Sec. II, Baker[19] proves that from the series expansion for finite box size with Dirichlet boundary conditions

$$f = f_-$$
$$= -\lim_{n\to\infty} \left[\max_L \left((1/L^4) \ln[\exp(\lambda_0 v N^4 Y) B_{n,-1}(\lambda)] \right) \right] , \tag{3.13}$$

where vN^4/L^4 is a pure number depending only on lattice structure, and f and f_- are as given by (3.6) and (3.7). If the series expansion for free-boundary conditions is used, then the result,

$$f_+ = -\lim_{n\to\infty} \left[\min_L \left((1/L^4) \ln[\exp(\lambda_0 v N^4 Y) B_{n,0}(\lambda)] \right) \right]$$
$$\leqslant f = f_- \tag{3.14}$$

follows, where the inequality follows by (3.2).

By means of the cluster property for the free field ($\lambda_0 \equiv 0$) special case and the combinatorial structure of the series expansion in λ_0 Baker[19] shows that the expansion is finite, term by term, and independent of L since, by construction (3.11),

$$f_{+,L} - (1/L^4) \ln[\exp(\lambda_0 v N^4 Y) B_{n,0}(\lambda_0)] = O(\lambda_0^{2n+1}) \,,$$
$$f_{-,L} - (1/L^4) \ln[\exp(\lambda_0 v N^4 Y) B_{n,0}(\lambda_0)] = O(\lambda_0^{2n}) \,, \tag{3.15}$$

and since all the series terms remain finite as $L\to\infty$, Eq. (3.15) remains valid in the infinite volume limit.

Next I remark that (3.4) can be replaced by a uniform bound S_n for $0 \leqslant \lambda_0 \leqslant \infty$. The end point $\lambda_0 = 0$ causes no trouble as it is explicitly computable and finite ($a > 0, m_0^2 > 0$). Since $:\mathscr{P}:$ is lower semibounded, the factor exp[ ] tends to a finite number of peaks at fixed values of x. By the saddle-point method, as $\lambda_0 \to \infty$ the bound in (3.3) becomes

$$\frac{\Sigma_i x_i^n \alpha_i}{\Sigma_i \exp(-8a^2 x_i^2) \alpha_i} \,, \tag{3.16}$$

where only the tallest peaks contribute, and $\Sigma \alpha_i = 1$. Thus the upper bound of (3.3) is uniformly bounded over the whole range $0 \leqslant \lambda_0 \leqslant \infty$ and we select S_n as the least upper bound. As a consequence of this bound, it is a straightforward computation to show that $|f'_-|$ and $|f'|$ are uniformly bounded for all L and $0 \leqslant \lambda_0 \leqslant \infty$ if $a > 0$, $m_0^2 > 0$. Thus it must be that the infinite volume limit of $f_- = f$ is a uniformly continuous function of λ_0, $0 \leqslant \lambda_0 \leqslant \infty$.

If we investigate the behavior of $f_\pm$ as a function of the boundary conditions we note that the difference between Z_+ and Z_- is just a change in the coefficients of the boundary $\phi_{\mathbf{r}}^2$ terms. Thus if we go from Z_+ to Z_- by adding $\eta\, va^{-2}\phi_{\mathbf{r}}^2$, η running from zero to unity, to each boundary $\phi_{\mathbf{r}}^2$ term we find

$$\frac{\partial f}{\partial \eta} = \frac{va^{-2}}{L^4} \sum_{\partial} \langle \phi_{\mathbf{r}}^2 \rangle \,. \tag{3.17}$$

By (3.3) each term is bounded and as there are only of order L^3 terms $\partial f/\partial\eta \to 0$ as $L\to\infty$. Thus $f_+ = f_- = f$ in the infinite volume limit.

If is now established that, for any of the three boundary conditions considered, the same infinite volume limits of the free energy per unit volume is obtained. For Dirichlet boundary conditions the infinite volume limit of the Schwinger functions exists. For the approach to the origin along the positive, real λ_0 axis, these functions are asymptotic to the finite, term-by-term, perturbation series in λ_0. Uniform continuity holds for the free energy in λ_0 and also the other coefficients of $\not{h}$ and the infinite volume limits of the Schwinger functions are uniformly bounded in λ_0.

I remind the reader that so far I have only discussed existence and have not discussed the question of uniqueness of the sum of even the finite volume the perturbation series. A partial discussion of the direct summation of the infinite volume limit perturbation series will be given in Sec. V.

To prepare for the eventual consideration of the continuum limit I now consider the properties of the two-point Schwinger function $S(\mathbf{r}_1,\mathbf{r}_2)$. As is well known (e.g., Baker[19]), for a lattice free field [Eq. (2.2), $\lambda_0 \equiv 0$]

$$S(\mathbf{r},\mathbf{s}) \propto e^{-m_0|\mathbf{r}-\mathbf{s}|}/|\mathbf{r}-\mathbf{s}|^{3/2} . \tag{3.18}$$

This exponential decay identifies m_0 as the mass of the field and this property is useful in making further progress on the rigorous properties of the model field theory. For the lattice cutoff Euclidean theory under discussion we wish the behavior

$$\langle \phi_0 \phi_\mathbf{r} \rangle \propto \exp(-m(\hat{\mathbf{r}})|\mathbf{r}|) , \tag{3.19}$$

where $\hat{\mathbf{r}}$ is a unit vector parallel to $\mathbf{r}$. We must allow possible direction dependence at this stage since we are working on a lattice.

The usual definition of the physical mass in the infinite volume limit is

$$m = \liminf_{|\mathbf{r}| \to \infty} (-\ln[\langle \phi_0 \phi_\mathbf{r} \rangle / B]/|\mathbf{r}|) . \tag{3.20}$$

By Griffiths' inequalities,[22,23] which hold for this case, $\langle \phi_0 \phi_\mathbf{r} \rangle \geqslant 0$ so there is no problem taking the logarithm, and B is the upper bound of (3.3) for $\langle \phi_0 \phi_\mathbf{r} \rangle$ uniform over all $\mathbf{r}$. By construction $m \geqslant 0$. It is worth mentioning that a finite box size approximation to m can be defined. Instead of (3.19), define

$$\mu(L) = \min_{\mathbf{r},\mathbf{s}} (-\ln[\langle \phi_\mathbf{r} \phi_\mathbf{s} \rangle / B]/|\mathbf{r}-\mathbf{s}|) . \tag{3.21}$$

By arguments of Baker[19] it follows directly that

$$\mu(L+1) \leqslant \mu(L) , \tag{3.22}$$

and by construction $\mu(L) \geqslant 0$. Thus

$$\lim_{L \to \infty} \mu(L) = \mu \geqslant 0 , \tag{3.23}$$

by standard theorems. Next, I need a result of Nelson's.[16] First the unnormalized transfer matrix from one lattice hyperplane to the next is

$$\begin{aligned} \mathsf{T}_0 = \exp\Bigg(-\frac{1}{2} v \sum_{n=0}^{1} \sum_{\mathbf{r}}{}' \Bigg\{ & \sum_{\{\boldsymbol{\delta}\}'} [(\phi_{\mathbf{r}+\boldsymbol{\delta}+n\boldsymbol{\delta}'} - \phi_{\mathbf{r}+n\boldsymbol{\delta}'})^2 a^{-2}] \\ & + m_0^2 \phi_{\mathbf{r}+n\boldsymbol{\delta}'}^2 + \lambda_0 {:}\not{h}(\phi_{\mathbf{r}+n\boldsymbol{\delta}'}){:} \Bigg\} \\ & + v \sum_{\mathbf{r}}{}' (\phi_{\mathbf{r}+\boldsymbol{\delta}'} - \phi_\mathbf{r})^2 a^{-2} \Bigg) , \end{aligned} \tag{3.24}$$

where $\Sigma_\mathbf{r}{}'$ is over the three-dimension lattice hyperplane, $\{\boldsymbol{\delta}\}'$ is the subset of $\{\boldsymbol{\delta}\}$ that lies in the lattice hyperplane, and $\boldsymbol{\delta}'$ is the nearest-neighbor vector perpendicular to the hyperplane. The normalized transfer matrix is

$$\mathsf{T} = \mathsf{T}_0/\mathrm{spr}(\mathsf{T}_0) , \tag{3.25}$$

where the spectral radius of an operator is defined as (independent of norm)

$$\mathrm{spr}(\mathsf{A}) = \lim_{n \to \infty} \|\mathsf{A}^n\|^{1/n} . \tag{3.26}$$

It is evident by construction that T is a real symmetric operator and positively preserving, i.e., if u is everywhere positive, then $\mathsf{T}u$ is also. Further all eigenvalues, t_ν, of T must by standard theorems be real, and by construction $|t_\nu| \leqslant 1$. In fact Nelson[16] shows that they are all non-negative and therefore, there exists a unique, Hermitian, positive operator H such that

$$\mathsf{T} = e^{-\mathsf{H}} . \tag{3.27}$$

A consequence of this result is, by the spectral theorem for T,

$$\frac{\langle \phi_0 \phi_{n\boldsymbol{\delta}'} \rangle}{\langle \phi_0^2 \rangle} = \frac{\langle \phi_0 e^{-n\mathsf{H}} \phi_0 \rangle}{\langle \phi_0^2 \rangle} = \int_0^1 t^n \, d\mu(t) , \tag{3.28}$$

where $d\mu$ is non-negative. Since $\langle t^{2n} \rangle \geqslant \langle t^n \rangle^2$ we find that, in any direction perpendicular to a lattice hyperplane (or any reflection symmetry hyperplane of the lattice) that

$$\frac{-\ln(\langle \phi_0 \phi_{n\boldsymbol{\delta}'} \rangle / \langle \phi_0^2 \rangle)}{|n\boldsymbol{\delta}'|} \geqslant \frac{-\ln(\langle \phi_0 \phi_{2n\boldsymbol{\delta}'} \rangle / \langle \phi_0^2 \rangle)}{|2n\boldsymbol{\delta}'|} . \tag{3.29}$$

In addition (3.28) has the obvious consequence (as $0 \leqslant t \leqslant 1$) that the correlations are monotonic along lattice lines:

$$\langle \phi_0 \phi_{n\boldsymbol{\delta}'} \rangle \leqslant \langle \phi_0 \phi_{(n-1)\boldsymbol{\delta}'} \rangle , \tag{3.30}$$

which implies $\langle \phi_0^2 \rangle$ bounds every two-point correlation along a lattice line.

The property of reflection positivity (or Osterwalder–Schrader positivity) holds for our model.[24–28] Combining this property with further arguments, Schrader[29] has extended (3.30) to show that $\langle \phi_0 \phi_\mathbf{r} \rangle$ is a monotonically decreasing function of each $|x_i|$ [$\mathbf{r} = (x_1,x_2,x_3,x_4)$] separately. This result means, of course,

$$0 \leqslant \langle \phi_\mathbf{r} \phi_\mathbf{s} \rangle \leqslant \langle \phi_0^2 \rangle . \tag{3.31}$$

When Schrader monotonicity is coupled with (3.29) we conclude that as the box size goes to infinity, the separation $|\mathbf{r}-\mathbf{s}|$ of the minimizing pair in (3.21) goes to infinity, and so μ defined by (3.23) and m defined by (3.20) agree with each other. (See the Appendix.)

Another consequence of reflection positivity is that if θ is a reflection in the lattice hyperplane and $F(\phi_1,\ldots,\phi_n)$ and $G(\phi_1,\ldots,\phi_n)$ are functions of ϕ_i's lying all on one side of the hyperplane, then the standard reflection-positivity inequality,

$$\langle F\theta F^* \rangle \geqslant 0 , \tag{3.32}$$

where * denotes complex conjugate, can be used to prove the Frölich–Lieb[30] inequality

$$|\langle F\theta G^* \rangle|^2 \leqslant \langle F\theta F^* \rangle \langle G\theta G^* \rangle , \tag{3.33}$$

which in turn leads to[31,32]

$$|\langle\phi^A\phi^B\rangle - \langle\phi^A\rangle\langle\phi^B\rangle| = |\langle(\phi^A - \langle\phi^A\rangle)(\phi^B - \langle\phi^B\rangle)\rangle|$$
$$\leqslant \langle\phi^A\theta\phi^A\rangle^{1/2}\langle\phi^B\theta\phi^B\rangle^{1/2}$$
$$\leqslant \langle(\phi^2)^A\rangle^{1/2}\langle\phi^B\theta\phi^B\rangle^{1/2}, \qquad (3.34)$$

where A and B are on opposite sides of the hyperplane and the last inequality follows in a manner similar to (3.30) *et seq.* By ϕ^A we mean $\phi^A = \Pi_{a\in A}\phi_a$, where A is a set of lattice sites perhaps including repeats.

Another valuable tool in the study of this model is the Fortuin, Kasteleyn, and Ginibre[31] inequalities, which has been proved for this case. They state that if $f(\phi_1,...,\phi_n)$ and $g(\phi_1,...,\phi_n)$ are increasing functions of each argument separately, then

$$\langle fg\rangle \geqslant \langle f\rangle\langle g\rangle . \qquad (3.35)$$

By use of these inequalities following the line of argument of Bricmont *et al.*[32] and using superstable estimates[33,34] Sokal[35] has established a cluster property. Superstable means that the condition

$$U(\phi_\Lambda) \geqslant \sum_{\lambda\in\Lambda} A\phi_\lambda^2 - C \qquad (3.36)$$

is required, where Λ is the finite portion of the infinite space lattice considered, $A > 0$, C is real and the partition function is given by

$$Z_\Lambda = \int \exp[-U(\phi_\Lambda)] \prod_{\lambda\in\Lambda} d\mu(\phi_\lambda) . \qquad (3.37)$$

This condition can be met easily for (2.2) by adjusting the ϕ^2 term, which goes with the site–site interaction, and placing the rest in the $d\mu$ term. The Sokal cluster property is

$$|\langle\phi^A\phi^B\rangle - \langle\phi^A\rangle\langle\phi^B\rangle| \leqslant C|\ln x|^{\alpha+\beta-2}x, \qquad (3.38)$$

where C is a constant, α is the number of members of A and β of B, and

$$x = \sum_{a\in A}\sum_{b\in B} \langle\phi_a\phi_b\rangle - \langle\phi_a\rangle\langle\phi_b\rangle . \qquad (3.39)$$

Note that $\langle\phi_a\rangle = \langle\phi_b\rangle = 0$ by the $\phi\to-\phi$ symmetry of the model in the absence of spontaneous symmetry breaking. Clearly (3.39) implies that if the two-point correlation function decays exponentially, then so too do the multipoint correlation functions except perhaps for a power of the separation.

Suppose now that we have chosen the parameters of (2.2) in such a manner that $m > 0$ (3.20). Let us compute

$$\frac{\partial S(\mathbf{r}_1,...,\mathbf{r}_n)}{\partial\lambda_0} = \sum_{\mathbf{t}} \langle\phi_{\mathbf{r}_1},...,\phi_{\mathbf{r}_n}\not{h}(\phi_{\mathbf{t}})\rangle - \langle\phi_{\mathbf{r}_1},...,\phi_{\mathbf{r}_n}\rangle\langle\not{h}(\phi_{\mathbf{t}})\rangle . \qquad (3.40)$$

The above sum over t breaks up into two parts. In the first part there are a finite number of lattice sites $\mathbf{t}$ that cannot be separated from $\{\mathbf{r}_1,...,\mathbf{r}_n\}$ by a hyperplane. In the second part (remainder) this separation can be made. For the second part, by (3.34),

$$\left|\sum_{\mathbf{t}}^{(2)} \langle\phi_{\mathbf{r}_1},...,\phi_{\mathbf{r}_n}\not{h}(\phi_{\mathbf{t}})\rangle - \langle\phi_{\mathbf{r}_1},...,\phi_{\mathbf{r}_n}\rangle\langle\not{h}(\phi_{\mathbf{t}})\rangle\right|$$
$$\leqslant \sum_{\mathbf{t}} \langle\phi_{\mathbf{r}_1}^2,...,\phi_{\mathbf{r}_n}^2\rangle^{1/2}\langle\not{h}(\phi_{\mathbf{t}})\theta_{\mathbf{t}}\not{h}(\phi_{\mathbf{t}})\rangle^{1/2}, \qquad (3.41)$$

where $\theta_{\mathbf{t}}$ is a reflection through that lattice hyperplane which separates $\{\mathbf{r}_1,...,\mathbf{r}_n\}$ from $\mathbf{t}$ and maximizes $|\mathbf{t} - \theta\,\mathbf{t}|$. By the assumed exponential decay, (3.3) and (3.38), the sum over the second part is finite. As there are only a finite number of terms in the first part it is also finite. Thus $\partial S/\partial\lambda_0$ is finite, independent of L, and so also as $L\to\infty$. Thus we may conclude that $\mathscr{S}_-$ is uniformly continuous in λ_0. The argument is the same for the other parameters in $\not{h}$.

Again, suppose that $m > 0$, following the analysis at (3.17) we find that, for the difference between free and Dirichlet boundary conditions,

$$\frac{\partial S(\mathbf{r}_1,...,\mathbf{r}_n)}{\partial\eta} = va^{-2}\sum_{\partial} \{\langle\phi_{\mathbf{r}_1},...,\phi_{\mathbf{r}_n}\phi_{\mathbf{t}}^2\rangle - \langle\phi_{\mathbf{r}_1},...,\phi_{\mathbf{r}_n}\rangle\langle\phi_{\mathbf{t}}^2\rangle\} . \qquad (3.42)$$

By the application of (3.38) the right-hand side is of the order $L^{n+3}\exp(-mL)$, which goes to zero as $L\to\infty$. Thus in the infinite volume limit $S(\mathbf{r}_1,...,\mathbf{r}_n)$ is independent of the boundary conditions.

For the next discussion I continue to assume that m (3.20) is greater than zero. The purpose of this discussion is to establish the continuity of the physical mass and amplitude renormalization constants as a function of the parameters of the Hamiltonian. Instead of the true mass, I will use the second moment definition of the mass. Namely,

$$(m_2)^2 = \frac{8[\Sigma_{\mathbf{r}}\langle\phi_0\phi_{\mathbf{r}}\rangle]}{[\Sigma_{\mathbf{r}}\mathbf{r}^2\langle\phi_0\phi_{\mathbf{r}}\rangle]} . \qquad (3.43)$$

Since $\langle\phi_0\phi_{\mathbf{r}}\rangle \geqslant 0$ by the Griffiths inequality and at least one term in the denominator is strictly positive for (2.2) Eq. (3.43) is well defined, provided $m > 0$ to assure the convergence of the sums. Since we have shown at (3.42) that $\langle\phi_0\phi_{\mathbf{r}}\rangle$ is a continuous function of λ_0, given $m > 0$, then any partial sum of $\Sigma_{\mathbf{r}}\mathbf{r}^2\langle\phi_0\phi_{\mathbf{r}}\rangle$ and $\Sigma_{\mathbf{r}}\langle\phi_0\phi_{\mathbf{r}}\rangle$ is continuous as the sum of a finite number of continuous functions is continuous. For any closed interval I_1, $\lambda_1 \leqslant \lambda_0 \leqslant \lambda_2$, in which $m > 0$ for all λ_0, each partial sum is, of course, uniformly continuous. By the exponential decay of $\langle\phi_0\phi_{\mathbf{r}}\rangle \geqslant 0$ the partial sums from (3.43) converge uniformly over I_1 and so the complete sum is by standard theorems, also continuous and thus the limiting function m_2 is also continuous in I_1.

To complete the control of the two-point function, the usual procedure is to scale the amplitude of the field $\phi_{\mathbf{r}}$ as

$$\psi_{\mathbf{r}} = Z_3^{-1/2}\phi_{\mathbf{r}} , \qquad (3.44)$$

where Z_3 is the traditional name for this factor. It is determined by the equation

$$v\sum_{\mathbf{r}} \langle\psi_0\psi_{\mathbf{r}}\rangle = m^{-2} , \qquad (3.45)$$

so the result is

$$Z_3 = m^2 v \sum_{\mathbf{r}} \langle\phi_0\phi_{\mathbf{r}}\rangle , \qquad (3.46)$$

or m_2 is also frequently used in place of m in (3.46). Needless to say, in the regions where $m = 0$ it is of course, continuous in λ_0, etc. The above results do not address the question of whether m can drop discontinuously to zero or not, nor will I in this paper.

Next I show that at least for some sets of parameters, m

is greater than zero. It is most convenient to use the formulation (2.5) for this demonstration. First, change variables to

$$\sigma_{\mathbf{r}} = \nu_{\mathbf{r}} s_{\mathbf{r}}, \quad \nu = \pm 1, \quad 0 \leqslant s \leqslant \infty, \tag{3.47}$$

so that (2.5) can be rewritten as

$$\begin{aligned} Z = M^{-1} & \sum_{\{\nu_{\mathbf{r}} = \pm 1\}} \int_0^{\infty} \cdots \int \prod ds_{\mathbf{r}} \\ & \times \exp\Big\{ K \sum_{\mathbf{r}} \sum_{\{\boldsymbol{\delta}\}} \nu_{\mathbf{r}} n_{\mathbf{r}+\boldsymbol{\delta}} s_{\mathbf{r}} s_{\mathbf{r}+\boldsymbol{\delta}} \\ & - \sum_{\mathbf{r}} [\tilde{A} s_{\mathbf{r}}^2 + \tilde{\lambda}_0 P(s_{\mathbf{r}})] \Big\}, \end{aligned} \tag{3.48}$$

as P is an even polynomial. Now if we use the standard method,

$$\begin{aligned} & \exp\{K \nu_{\mathbf{r}} \nu_{\mathbf{r}+\boldsymbol{\delta}} s_{\mathbf{r}} s_{\mathbf{r}+\boldsymbol{\delta}}\} \\ & = \cosh[K s_{\mathbf{r}} s_{\mathbf{r}+\boldsymbol{\delta}}](1 + v_{\mathbf{r},\mathbf{r}+\boldsymbol{\delta}} \nu_{\mathbf{r}} \nu_{\mathbf{r}+\boldsymbol{\delta}}), \end{aligned} \tag{3.49}$$

where

$$v_{\mathbf{r},\mathbf{r}+\boldsymbol{\delta}} = \tanh[K s_{\mathbf{r}} s_{\mathbf{r}+\boldsymbol{\delta}}], \tag{3.50}$$

then Fisher[36] has proved that, before the integration over $s_{\mathbf{r}}$ is performed,

$$\langle \sigma_{\mathbf{r}} \sigma_{\mathbf{s}} \rangle \leqslant C(\mathbf{r},\mathbf{s}), \tag{3.51}$$

where $C(\mathbf{r},\mathbf{s})$ is the generating function for all self-avoiding random walks between $\mathbf{r}$ and $\mathbf{s}$ on the lattice. That is,

$$C(\mathbf{r},\mathbf{s}) = \sum_{\Gamma(\mathbf{r},\mathbf{s})} \prod_{(\mathbf{t},\mathbf{u}) \in \Gamma} v_{\mathbf{t},\mathbf{u}}. \tag{3.52}$$

Now to bound (3.52) after the integration, I first observe the elementary results, $x, y \geqslant 0$,

$$\begin{aligned} & \tanh x \leqslant x, \\ & 1 \leqslant \cosh Kxy \leqslant e^{Kxy} \leqslant \exp[\tfrac{1}{2} K(x^2 + y^2)]. \end{aligned} \tag{3.53}$$

It is sufficient to use the lower bound of (3.53) on every bond attached to a vertex in the random walk in the denominator of the expectation value over the $s_{\mathbf{r}}$ and the upper bound on the same bonds in the numerator. This procedure splits the total integral into two factors, one over the vertices of the random walk and the other over all other vertices. The latter cancels between numerator and denominator. The contribution of each vertex is then bounded by the finite factor,

$$V = \frac{\int_0^\infty x^2 \exp[\frac{1}{2} qKx^2 - \tilde{A} x^2 - \tilde{\lambda}_0 P(x)]\, dx}{\int_0^\infty \exp[-\tilde{A} x^2 - \tilde{\lambda}_0 P(x)]\, dx}, \tag{3.54}$$

where q is the lattice coordination number. Thus

$$C(\mathbf{r},\mathbf{s}) \leqslant V \sum_{\Gamma(\mathbf{r},\mathbf{s})} (KV)^{n(\Gamma)} \equiv \tilde{C}(\tilde{\mathbf{r}},\tilde{\mathbf{s}}), \tag{3.55}$$

the sum is over all self-avoiding walks Γ, and where $n(\Gamma)$ is the number of bonds in the walk Γ. Next, a weaker bound will result if the self-avoiding restriction is dropped. Thus

$$\tilde{C}(\mathbf{r},\mathbf{s}) \leqslant V \sum_{\Gamma(\mathbf{r},\mathbf{s})} (KV)^{n(\Gamma)} \equiv \hat{C}(\mathbf{r},\mathbf{s}), \tag{3.56}$$

where the sum is now over any walk which connects $\mathbf{r}$ and $\mathbf{s}$. The sum defining $\hat{C}$ has been evaluated. By translation invariance it suffices to consider

$$\hat{C}(0,\mathbf{s}) = \frac{V}{(2\pi)^4} \int_{-\pi}^{\pi} \cdots \int \frac{\exp(ia^{-1}\mathbf{s}\cdot\mathbf{k})\, d^4\boldsymbol{\theta}}{1 - qVK\lambda(\boldsymbol{\theta})}, \tag{3.57}$$

where q is the lattice coordination number, and $\lambda(\boldsymbol{\theta})$ is the lattice structure function. It is defined by

$$\begin{aligned} \lambda(\mathbf{k}) & \equiv q^{-1} \sum_{\mathbf{r}} \exp(ia^{-1}\mathbf{r}\cdot\mathbf{k}) \\ & \simeq 1 - \tfrac{1}{2} k^2/d + O(r^4 k^4), \end{aligned} \tag{3.58}$$

where the sum over $\mathbf{r}$ is over the set of nearest-neighbor lattice vectors. The asymptotic behavior of (3.57) as $|\mathbf{s}| \to \infty$ has been computed by Montroll and Weiss.[37] It is

$$\begin{aligned} \hat{C}(0,\mathbf{s}) & \sim \frac{2V}{(\frac{1}{2}\pi qVK)^2} \left(\frac{a}{s}\right) [qVK(1 - qVK)]^{1/2} \\ & \times K_1 \left(\frac{2s}{a} \left(\frac{1 - qVK}{qVK} \right)^{1/2} \right) \sim \frac{\hat{m}^2}{qK} \left(\frac{2}{\pi \hat{m} s} \right)^{3/2} e^{-\hat{m}s}, \end{aligned} \tag{3.59}$$

where K_1 is the first modified Bessel function of the second kind and

$$\hat{m} = (2/a)[(1 - qVK)/(qVK)]^{1/2}. \tag{3.60}$$

These results are valid for $0 < qVK < 1$. From (3.54) it is clear that $KV(K)$ is an unbounded, monotonically increasing function of K so there exists a $\hat{K}$ such that $0 < K < \hat{K}$ exactly corresponds to the aforementioned range for qVK. Thus from (3.59), the definition of the physical mass (3.20), and the inequalities (3.51), (3.55), and (3.56), I conclude that

$$m \geqslant \hat{m}. \tag{3.61}$$

By (3.60), and the remark that $K = 0$ means $\langle \sigma_{\mathbf{r}} \sigma_{\mathbf{s}} \rangle = \delta_{\mathbf{r},\mathbf{s}}$ so $m = \infty$, it now follows that for $0 \leqslant K < \hat{K}$, $m > 0$. Thus at least in this region the above given analysis holds.

As a further result, I show that for K sufficiently small, the series expansion for f converges. To start with I rewrite (2.5) as

$$\begin{aligned} Z & = M^{-1} \int_{-\infty}^{+\infty} \cdots \int \prod_{\mathbf{r}} d\sigma_{\mathbf{r}} \exp\Big\{ -\frac{1}{2} K \sum_{\mathbf{r}} \sum_{\{\boldsymbol{\delta}\}} (\sigma_{\mathbf{r}} - \sigma_{\mathbf{r}+\boldsymbol{\delta}})^2 \\ & \quad - \sum_{\mathbf{r}} \Big[\Big(\tilde{A} - \frac{1}{2} qK \Big) \sigma_{\mathbf{r}}^2 + \tilde{\lambda}_0 P(\sigma_{\mathbf{r}}) \Big] \Big\} \\ & = M^{-1} \int_{-\infty}^{+\infty} \cdots \int \prod_{\mathbf{r}} d\sigma_{\mathbf{r}} \prod_{\mathbf{r},\{\boldsymbol{\delta}\}} (1 + f_{\mathbf{r},\mathbf{r}+\boldsymbol{\delta}}) \\ & \quad \times \exp\Big\{ - \sum_{\mathbf{r}} \Big[\Big(\tilde{A} - \frac{1}{2} qK \Big) \sigma_{\mathbf{r}}^2 + \tilde{\lambda}_0 P(\sigma_{\mathbf{r}}) \Big] \Big\} \\ & = \sum_{\Gamma} \int_{-\infty}^{+\infty} \cdots \int \prod_{\mathbf{r}} d\sigma_{\mathbf{r}} \prod_{\mathbf{r},\{\boldsymbol{\delta}\} \in \Gamma} f_{\mathbf{r},\mathbf{r}+\boldsymbol{\delta}} \\ & \quad \times \exp\Big\{ - \sum_{\mathbf{r}} \Big[\Big(\tilde{A} - \frac{1}{2} qK \Big) \sigma_{\mathbf{r}}^2 + \tilde{\lambda}_0 P(\sigma_{\mathbf{r}}) \Big] \Big\} \\ & = \sum_{\Gamma} \Big\langle \prod_{\mathbf{r},\{\boldsymbol{\delta}\} \in \Gamma} f_{\mathbf{r},\mathbf{r}+\boldsymbol{\delta}} \Big\rangle, \end{aligned} \tag{3.62}$$

where the sum over Γ is the sum over all subsets of the nearest-neighbor bonds on the finite portion of the lattice under consideration and

$$f_{\mathbf{r},\mathbf{r}+\boldsymbol{\delta}} = \exp\{-\tfrac{1}{2} K(\sigma_{\mathbf{r}} - \sigma_{\mathbf{r}+\boldsymbol{\delta}})^2\} - 1. \tag{3.63}$$

By the well-known combinatorial cluster theorem,

$$Z = \exp\Big\{ \sum_{\Gamma} \Big\langle \prod_{\mathbf{r},\mathbf{r}+\boldsymbol{\delta} \in \Gamma} f_{\mathbf{r},\mathbf{r}+\boldsymbol{\delta}} \Big\rangle \Big\}, \tag{3.64}$$

where the sum over Γ is now restricted to connected clusters only. If we organize the sum by the number of vertices, then by adapting an argument of Glimm and Jaffee[38] to four dimensions there are at most $(2q^{2q+1})^v$ different connected terms, for v vertices. Next since

$$\begin{aligned}|e^{-x}-1| &\leqslant \left|x-\frac{x^2}{2!}+\frac{x^3}{3!}-\frac{x^4}{4!}+\cdots\right| \\ &\leqslant |x|+\frac{|x|^2}{2!}+\frac{|x|^3}{3!}+\frac{|x|^4}{4!}+\cdots \\ &\leqslant |x|+\frac{|x|^2}{1!}+\frac{|x|^3}{2!}+\frac{|x|^4}{3!}+\cdots \\ &= |x|e^{|x|}, \end{aligned} \tag{3.65}$$

and

$$\exp[\tfrac{1}{2}|K|(\sigma_r-\sigma_{r+\delta})^2]\leqslant\exp[|K|(\sigma_r^2+\sigma_{r+\delta}^2)], \tag{3.66}$$

it follows that

$$|f_{r,r+\delta}|\leqslant|K|(\sigma_r^2+\sigma_{r+\delta}^2)\exp[|K|(\sigma_r^2+\sigma_{r+\delta}^2)]. \tag{3.67}$$

Thus the contribution to (3.64) from v vertex clusters is bounded by

$$\begin{aligned} N(2q^{2q+1})^v &\int_{-\infty}^{+\infty}\cdots\int \prod d\sigma_r \prod |f_{r,r+\delta}| \\ &\times\exp\left\{-\sum_r\left[\left(\tilde{A}-\frac{1}{2}q|K|\right)\sigma_r^2+\tilde{\lambda}_0P(\sigma_r)\right]\right\} \\ &\times\left|\int_{-\infty}^{+\infty}d\sigma_r\,\exp\left\{-\left(\tilde{A}-\frac{1}{2}qK\right)\sigma_r^2+\tilde{\lambda}_0P(\sigma_r)\right\}\right|^{-v}, \end{aligned} \tag{3.68}$$

where the configuration in the numerator is that which makes the maximum contribution. Since no more than q bonds can meet at any one site, it follows that for K sufficiently small, independent of the number of vertices, the largest configuration will be the one with the fewest factors of K, i.e., $v-1$. Thus the numerator integral is bounded by

$$\begin{aligned} |K|^{v-1}&\left|\int_{-\infty}^{+\infty}d\sigma^{2q}[\mathrm{Max}(1,\sigma^{2q})]\right. \\ &\left.\times\exp\left\{-\left(TA-\frac{1}{2}qK-q|K|\right)\sigma_r^2-T\lambda_0P(\sigma_r)\right\}\right|^v, \end{aligned} \tag{3.69}$$

where use of the fact that $\langle\sigma^{2r+2s}\rangle\geqslant\langle\sigma^{2r}\rangle\langle\sigma^{2s}\rangle$ was made. Thus the contribution of v vertex connected clusters to (3.64) is bounded by

$$(N/K)(2q^{2q+1}|K|F(K,|K|))^v, \tag{3.70}$$

which, by direct computation, goes to zero geometrically with v as $v\to\infty$ for some $|K|>0$ small enough. Thus, as each term in the sum in (3.64) is analytic, and it is absolutely convergent for small enough $|K|$, the sum itself is analytic at least in the same region by the standard theorems of complex variable theory and so the series expansion in K is a convergent one. The same arguments hold if a sum over a finite number of sites, $\Sigma_{r\in A}\,\sigma_rH_r$, is added to the exponents in (2.5). If we differentiate with respect to all the H_r, $r\in A$, then the only terms that remain are connected clusters B such that $A\subset B$, and the above argument implies the convergence of the K series for $|K|$ small enough for the Schwinger functions as well as the free energy per unit volume.

In addition, as a final result in this section I remark that the bound in (3.68) can be extended to complex $\tilde{\lambda}_0$, provided we hold fixed $\tilde{A}>0$ and do not impose (2.7). This case can be seen to be appropriate for the normal perturbation theory expansion. Clearly, for $|K|$ sufficiently small the $\exp(-\tilde{A}\sigma_r^2)$ factor alone suffices to make the integrals of (3.68) converge so that we may select any $\tilde{\lambda}_0$ we please with $\mathrm{Re}(\tilde{\lambda}_0)\geqslant 0$ and maintain convergence. If $|\tilde{\lambda}_0|\leqslant\lambda_M$ for some definite $\lambda_M>0$, then the denominator is well bounded away from zero. If I now gradually rotate the contour of integration over σ_r by an angle θ, $-\pi/4<\theta<\pi/4$, and P is of degree $2p$, we can extend analytically the bound (3.68) from $|\arg\tilde{\lambda}_0|\leqslant\pi/2$ without contour rotation to $|\arg\tilde{\lambda}_0|<(p+1)\pi/2$ by the union of a finite number of overlapping functional elements. Consequently, it follows that there exists a K_0, $|K_0|>0$ such that for $|K|<|K_0|$, the sum in (3.64) converges absolutely for $\tilde{\lambda}_0$ in the angular wedge $|\tilde{\lambda}_0|<\lambda_M$, $|\arg\tilde{\lambda}_0|<(p+1)\pi/2$ and so is analytic, in $\tilde{\lambda}_0$ in compact subsets of this wedge on the Riemann surface. The same type of argument, where $\tilde{\lambda}_0$ is real and positive, also establishes directly that for $|K|>0$ sufficiently small, there exists a neighborhood of $\tilde{\lambda}_0$, where f and the Schwinger functions are analytic in $\tilde{\lambda}_0$. These results will be useful in Sec. V when we discuss the summability of the perturbation series.

IV. CONTINUUM LIMIT

In the previous section I established the existence on a space lattice with a finite lattice spacing of the infinite volume or thermodynamic limit of certain boson, Euclidean polynomial quantum field theories. A number of properties of these limiting field theories were established. In this section the limit as the lattice spacing goes to zero, in such a way as to maintain control of the two-point Schwinger function, is discussed. The two main items to control are the rate of exponential decay m [Eq. (3.20)] and the amplitude renormalization factor Z_3 [Eq. (3.46)]. It will be most convenient to work with form (2.5) with normalization (2.7) of the partition function because all the parameters appearing are then of order unity with respect to the lattice spacing. It is convenient further to introduce the notation

$$\chi=\sum_r\langle\sigma_0\sigma_r\rangle,\quad \xi^2=\frac{\Sigma_r(r/a)^2\langle\sigma_0\sigma_r\rangle}{8\chi}, \tag{4.1}$$

which are the corresponding (unsubtracted) magnetic susceptibility χ and correlation length in the analogous, continuous-spin Ising model problem. As long as $m>0$ the sums in (4.1) must converge. The correlation length ξ is the dimensionless second moment definition measured in terms of the number of lattice spacings. By (2.6) and (3.43),

$$(m_2)^2a^2\xi^2=1 \tag{4.2}$$

is an identity, and if we replace m by m_2 in (3.45) we get a slightly modified amplitude renormalization constant by (2.6) and (3.46),

$$Z_3 = qK\chi/16\xi^2. \tag{4.3}$$

In order to take the appropriate continuum limit $a \to 0$, with m_2 fixed, we must, by (4.2), take $\xi^2 \to \infty$. This procedure is equivalent, on a fixed lattice-spacing lattice, to taking m_2, or m for that matter, to zero. We have shown for K small enough by (3.60) and (3.61) that $m \to \infty$ as $K \to 0$ and is definitely positive for $KV < q^{-1}$. By use of a Peierls-type argument[39–41] van Beijeren and Sylvester[42] have proved in two or higher dimensions that there is long-range order ($\langle \sigma_0 \sigma_{\mathbf{r}} \rangle \to M^2 > 0$) for K sufficiently large. From this result it follows by (3.20) that $m = 0$, for K sufficiently large, and that $\chi = \xi^2 = \infty$ in (4.1). As $\mu(L)$ [Eq. (3.21)] converges to m as $L \to \infty$, we can, for K large enough and L finite, make $\mu(L)$ as small as we please. But for finite L, $\mu(L)$ is continuous in K because $\langle \phi_{\mathbf{r}} \phi_{\mathbf{s}} \rangle$ is a finite sum of continuous functions, and $\mu(L)$ is given by one particular such $(\mathbf{r},\mathbf{s})$. As K changes either $(\mathbf{r},\mathbf{s})$ remains the minimum pair, assuring continuity, or there is a crossover to another $(\mathbf{r},\mathbf{s})$. If there is a crossover, it occurs at the value of K where the value of $\mu(L)$ computed from the two pairs is the same, also implying continuity. The resulting function may, however, only be piecewise differentiable. In any such section of differentiability, we compute

$$\frac{\partial \mu(L)}{\partial K} = \sum_{\mathbf{t}} \sum_{\{\delta\}} \frac{\langle \sigma_{\mathbf{r}} \sigma_{\mathbf{s}} \sigma_{\mathbf{t}} \sigma_{\mathbf{t}+\delta} \rangle - \langle \sigma_{\mathbf{r}} \sigma_{\mathbf{s}} \rangle \langle \sigma_{\mathbf{t}}\, \sigma_{\mathbf{t}+\delta} \rangle}{\langle \sigma_{\mathbf{r}} \sigma_{\mathbf{s}} \rangle} \leqslant 0, \tag{4.4}$$

by the Griffiths–Kelly–Sherman inequality.[22,23,42] Thus as we have seen $\mu(L)$ runs continuously and monotonically from ∞ for $K = 0$ to 0, for $K = \infty$, for fixed $\tilde{\lambda}_0$. Hence by Bolzano's theorem, for any fixed given $\hat{m}$, there exists $K(L,\hat{m},\tilde{\lambda}_0,a)$ such that $\mu(L) = \hat{m}$. By use of (3.46), the amplitude renormalization factor $Z_3(L,\hat{m},\tilde{\lambda}_0,a)$ is also directly computable.

I now define the following limiting process. First, select a mass $\hat{m}$. Define a correlation length $\xi(L)$ in units of the lattice spacing by

$$\mu^2(L)a^2\xi^2(L) = 1 \tag{4.5}$$

analogous to (4.2). Second, select a sequence of correlations lengths $\hat{\xi}_j$ such that $\hat{\xi}_j \to \infty$ as $j \to \infty$ and a sequence of $T_j > 1$ such that $T_j \to \infty$ as $j \to \infty$. For each j choose a box of edge $L_j = T_j\, \hat{\xi}_j a_j$, where $a_j = (\hat{m}\hat{\xi}_j)^{-1}$. Then solve for

$$K_j = K(L_j,\hat{m},\tilde{\lambda}_0,a_j), \quad Z_3(j) = Z_3(L_j,\hat{m},\tilde{\lambda}_0,a_j), \tag{4.6}$$

as explained above, via the $\sigma \to \phi$ (2.6) and $\phi \to \psi$ (3.44) transformations, these parameters define a sequence of fields $\psi_{\mathbf{r}}(j)$. As $j \to \infty$, the limits $L \to \infty$ and $a \to 0$ are taken simultaneously. By construction, at every step mass renormalization $\mu(L) = \hat{m}$ and amplitude renormalization (3.45) is maintained. Since we have demonstrated in the Appendix that the pseudomass μ converges to the true mass m, the limiting field $\psi_{\mathbf{r}}$ has this preselected physical mass. In the absence of rotational invariance, this result is proved only along directions perpendicular to lattice reflection symmetry planes; however, $m_\perp \geqslant m \geqslant \frac{1}{2} m_\perp$ restricts the usual mass in this case. As all the properties, such as reflection positivity, the Griffiths inequalities, and the FKG inequalities, hold uniformly in j, they and their sequela as well, such as the cluster property, hold in the limit $j \to \infty$.

It is worth remarking that while the above given limiting process proves the existence of a fixed mass $m > 0$, and amplitude renormalized, Euclidean boson quantum field theory for any desired positive mass and any even polynomial interaction as described and restricted in Sec. II, it does *not* prove that $\mu(K,L)$ is uniformly continuous in K for all L. In Sec. III, it was shown that so long as $m > 0$, continuity results; however, it does not preclude the possibility that the $L \to \infty$ limit with fixed positive lattice spacing, $m(K)$, can drop discontinuously to zero from some finite value. As monotonicity still holds, there can be only one such drop. Nevertheless, the limiting process still, in this situation, constructs a field theory which satisfies all of Nelson's axioms, except perhaps rotational invariance. Hence we have demonstrated that the field theory exists in the sense that it is defined by a convergent limiting process.

In the subsequent discussion[18] of the question of whether the field theory so defined is trivial (i.e., no scattering) or nontrivial, numerical estimates based on finite length series expansions in K will be used. As I remarked at (3.70) these series have a nonzero radius of convergence. For that method to succeed, it is necessary that $m(K)$ drop continuously to zero. That question and the computation of the two-particle scattering amplitude in the continuum limit are the two principle questions to be addressed by those numerical estimates.

V. PERTURBATION THEORY

The most important question concerning the perturbation theory is, of course, whether it is useful in the sense that it, at least in principle, determines the physical theory. I have not yet been able to answer this question in full, but do have some partial results. Specifically, starting from the formulation (2.5), for sufficiently small $|K|$. I can prove that the (necessarily lattice cutoff) series is summable and uniquely determines the physically correct values of the free energy and the Schwinger functions.

First we observe that the general power counting arguments for these theories, "phantom fields" as described in Sec. II, are the same as in ϕ^4 theory.[2] The point is that when a ϕ^{2n} vertex in the diagramatic representation of the expansion in $\tilde{\lambda}_0$ about the free field appears, it creates $(n > 2)$ extra internal momentum integrations beyond those in a simple ϕ^4 theory. However, in the diagrammatic representation this diagram is multiplied by an extra factor of $a^{2(n-2)}$, as we saw in Eq. (2.6), which serves to exactly cancel the ultraviolet divergence as $a \to 0$. This fact leads to a superficial degree of ultraviolet divergence of $4 - E$, where E is the number of external lines.

What is required is a bound on the magnitude of the coefficients of the $\tilde{\lambda}_0$ expansion. To this end I first treat the case where $P(\sigma) = \sigma^{2p}$ and the lattice is the hyper-simple cubic, for ease of exposition. The expansion of the partition function is

$$\ln Z = \ln Z(\tilde{\lambda}_0 = 0) + \sum_{m=1}^{\infty} \frac{\tilde{\lambda}_0^m}{m!} \sum_{C_m} W(C_m) = \sum_{m=0}^{\infty} z_m \tilde{\lambda}_0^m, \tag{5.1}$$

where the sum over C_m is the sum over the labeled m-point connected graphs with $2p$ coordinated vertices, and $W(C_m)$ is

$$W(C_m) = (-1)^m \int_{-\pi/a}^{\pi/a} \cdots \int \prod_{i=1}^{mp} d\mathbf{k}_i \times \frac{1}{\tilde{A} - 4K + 2K \sum_{\tau=1}^{4} \sin^2(\frac{1}{2}\mathbf{k}\cdot\mathbf{e}_\tau)} \prod_{j=1}^{m} \delta\left(\sum_{l\in v_j} \mathbf{k}_l\right), \tag{5.2}$$

where the $\mathbf{e}_\tau$ are the four unit vectors in the directions of the crystal axes, and v_j are the m sets of momenta at each of the m vertices. The number of lines is $\frac{1}{2}$ (each line has two ends) times the number of vertices time the vertex coordination number, $2p$. The magnitude of $W(C_m)$ is bounded by

$$|W(C_m)| < \left[\frac{2\pi}{a(\tilde{A} - 4K)}\right]^{m(p-1)d}, \tag{5.3}$$

where the lower bound for the denominators $(\tilde{A} - 4K)$ was used. This bound is grossly large as the integrals over k are reduced for most cases by a factor of $O(a^{2mp})$ but it will suffice for our present needs. The total number of connected graphs of the class considered is less than the whole number of such graphs, which is $(2mp)!/\{2^m(mp)!\}$. If this bound is combined with (5.1) and (5.3) the coefficient z_m of $\tilde{\lambda}_0^m$ is bounded by

$$|z_m| < A_m = \frac{(2mp)!}{2^m m!(mp)!} \left[\frac{2\pi}{a(\tilde{A} - 4K)}\right]^{m(p-1)d}, \tag{5.4}$$

which is the order $[(p-1)m]!$ times a geometric factor. The addition of the lower-order (in σ) terms in $P(\sigma_r)$ add contributions less rapidly growing than $[(p-1)m]!$ and do not disturb significantly the argument.

Graffi *et al.*[43] give a generalized version of Watson's theorem, namely the following theorem.

Theorem: Let D be a sector of an n-sheeted Riemann surface defined as $0 < |z| < B$, $|\arg z| < \theta$, $\frac{1}{2}m\pi < \theta < \frac{3}{2}m\pi$, where $m \geqslant 1$ is an integer. Let D_1 be the sector $|\arg z| \leqslant \delta < \theta - \frac{1}{2}m\pi$ and $\bar{D}$ be the sector $0 < |z| < B$, $|\arg z| \leqslant \delta$. Then, given the formal power series $\Sigma_{n=0}^{\infty} a_n z^n$, suppose that (i) $f(z)$ is a function regular in D with the formal series an asymptotic series uniformly in D,

$$f(z) = \sum_{n=0}^{N} a_n z^n + R_N(z); \tag{5.5}$$

and (ii) there are σ, C so that

$$|a_n| < C\sigma^n (nm)!, \quad |R_N(z)| < C\sigma^{N+1}[m(N+1)]!|z|^{N+1}, \tag{5.6}$$

uniformly in D and N. Then the series $\Sigma_{n=0}^{\infty} a_n z^n$ is Borel summable to $F(z)$ in $\bar{D}$. That is

$$f(z) = \int_0^\infty e^{-t} F(zt^m)\,dt, \tag{5.7}$$

where

$$F(z) = \sum_{n=0}^{\infty} \alpha_n z^n, \quad \alpha_n = a_n/(mn)!, \tag{5.8}$$

and $F(z)$ can be continued to a function regular in D_1.

Note that the statement of this theorem implies uniqueness in $\bar{D}$.

In order to apply this theorem to our present case, I observe that the strong control of the remainder follows by differentiation of (3.62) $N+1$ times with respect to $\tilde{\lambda}_0$ and following the same line of argument to the end of Sec. III to bound this derivative. The bound on the remainder then follows by Taylor's theorem with remainder. Combining this observation with the deduction at the end of Sec. III that one can select $\theta = \frac{1}{2}\pi(p+1)$ and $m = p-1$ in the theorem, I conclude for $|K| < |K_0|$ and $|\tilde{\lambda}_0| < \lambda_m$, $|\arg \tilde{\lambda}_0| \leqslant \delta < \pi$ that the perturbation series in $\tilde{\lambda}_0$ is uniquely summable via (5.7) and (5.8) to the physically correct function. Since we also have shown at the end of Sec. III that f is analytic for all real positive $\tilde{\lambda}_0$, we may extend the definition of the free energy f by analytic continuation over this whole range and conclude that the perturbation series uniquely defines f, provided K is sufficiently small. The same arguments apply, with minor variations, to the Schwinger functions. Now for $\tilde{\lambda}_0$ real and positive we have shown at (3.55) that the free-energy Schwinger functions are analytic in K provided $KV < 1$. Thus by analytic continuation in K we can extend the results of the summation of the lattice cutoff perturbation series to this larger region. Put otherwise, the K series for fixed $\tilde{\lambda}_0$ can be constructed from the lattice cutoff series in $\tilde{\lambda}_0$. The converse, i.e., that the convergent K series implies the $\tilde{\lambda}_0$ expansion is evident by the analytic character in K (5.2) of each term in the $\tilde{\lambda}_0$ expansion. So long as the K series for real $\tilde{\lambda}_0$ can be analytically continued, we can define, in principle, from the cutoff $\tilde{\lambda}_0$ perturbation series the physical quantities. The mass and amplitude renormalizations are directly performable in this region as explained above. For $\tilde{\lambda}_0 \equiv 0$ we have the free-field case and the series expansions in K converge for all $0 \leqslant \xi < \infty$, which suffices to construct the free-field theory. As long as the integral in (5.7) converges absolutely (known for K small enough) for real positive $\tilde{\lambda}_0$, since the terms of the series expansion in $\tilde{\lambda}_0$ are analytic (5.2) in K, this construction provides directly the required analytic continuation to represent the physical, lattice cutoff theory.

Although it is true that for $\tilde{\lambda}_0 \equiv 0$, we have the free-field case and here the K series converges for all $0 \leqslant \xi < \infty$, which suffices to construct the field theory, it does not seem likely that we may simply reverse the limits summation of $\tilde{\lambda}_0$ series for fixed ξ and $\xi \to \infty$ directly in (5.7) because even in the mass and amplitude renormalized $\lambda_0{:}\phi^4{:}_4$ theory it is well known[2] that the limiting coefficients are not finite term by term. Furthermore, of course, it may happen that for $\tilde{\lambda}_0 > 0$ analyticity in it breaks down before ξ reaches infinity thereby limiting this approach. On the other hand, nothing that we know so far precludes the possibility that cases exist where the field theories we showed to exist in Sec. IV can be reached via analyticity in K, and in fact such examples have been reported by Baker and Johnson.[1] In such favorable cases, where the continuum field theory is an analytic function in $\tilde{\lambda}_0 > 0$ and the boundary value of an analytic function in K, it may be possible (e.g., Baker[44]) to design a single-limit, summation process to sum directly to the continuum limit from the cutoff perturbation series in $\tilde{\lambda}_0$. In the mean time, the equivalent K-series method for fixed $\tilde{\lambda}_0$ is available.

To illustrate the use of the K series, I will now calculate to leading order in $\tilde{\lambda}_0$ the two-particle and three-particle scattering amplitudes for the case where $P(\sigma_r)$ is of degree 6. Specifically (2.5) becomes

$$Z = M^{-1}\int_{-\infty}^{+\infty}\cdots\int\prod_r d\sigma_r \exp\Big[\sum_r\Big(K\sum_{\{\delta\}}\sigma_r\sigma_{r+\delta} - \tilde{A}\sigma_r^2 - \tilde{g}_0\sigma_r^2 - \tilde{\lambda}_0\sigma_r^6 + \tilde{H}_r\sigma_r\Big)\Big]. \quad (5.9)$$

As Baker and Kincaid[5] have shown, the dimensionless renormalized four-line coupling constant is given by

$$g = -\left(\frac{v}{a^4}\right)\frac{\partial^2\chi/\partial H^2}{\chi^2\xi^4}, \quad (5.10)$$

where χ and ξ^2 are defined by (4.1) and

$$\frac{\partial^2\chi}{\partial H^2}\bigg|_{H=0} = \sum_{r,t,u} U_4(\sigma_0,\sigma_r,\sigma_t,\sigma_u), \quad (5.11)$$

where U_4 is the four-point Ursell function[45] that is, when $H_r = 0$ so there is up–down spin symmetry, in the high-temperature region

$$U_4(\sigma_0,\sigma_r,\sigma_t,\sigma_u) = \langle\sigma_0\sigma_r\sigma_t\sigma_u\rangle - \langle\sigma_0\sigma_r\rangle\langle\sigma_t\sigma_r\rangle - \langle\sigma_0\sigma_t\rangle\langle\sigma_r\sigma_u\rangle - \langle\sigma_0\sigma_u\rangle\langle\sigma_r\sigma_t\rangle. \quad (5.12)$$

In order to study the three-particle scattering amplitude we begin with the zero-momentum scattering amplitude[12] (connected part)

$$\langle 000|000\rangle_c = Z_3^3\, G^{(6)}_{\text{trunc}}(0,0,0,0,0,0), \quad (5.13)$$

which we can reexpress, following Baker and Kincaid,[5] as

$$\langle 000|000\rangle_c = \left(\frac{v}{a^4}\right)^2\frac{m^{-2}(\partial^4\chi/\partial H^4)}{\chi^3\xi^8}, \quad (5.14)$$

and thus define the dimensionless, six-line coupling constant as

$$\lambda = -m^2\langle 000|000\rangle_c = -\left(\frac{v^2}{a^2}\right)\frac{\partial^4\chi/\partial H^4}{\chi^3\xi^8}, \quad (5.15)$$

where $\partial^4\chi/\partial H^4$ is the sum over the six-point Ursell function as in (5.11) for $\partial^2\chi/\partial H^2$.

This computation will lead to a first-order expansion in $\tilde{g}_0$ and $\tilde{\lambda}_0$ of (5.10) and (5.15). I employ the linked-cluster expansion method of Wortis.[46] This method expresses the series directly in terms of the cumulants of the single spin distribution as in (2.7). For $\tilde{g}_0 = \tilde{\lambda}_0 = 0$, (2.7) implies that $\tilde{A} = \frac{1}{2}$, if we set $\Delta\tilde{A} = \tilde{A} - \frac{1}{2}$, then it is elementary to compute that, to linear order,

$$\begin{aligned}
\langle\sigma^2\rangle &= 1 - 90\tilde{\lambda}_0 - 12\tilde{g}_0 - 2\Delta\tilde{A},\\
\langle\sigma^4\rangle &= 3 - 900\tilde{\lambda}_0 - 96\tilde{g}_0 - 12\Delta\tilde{A},\\
\langle\sigma^6\rangle &= 15 - 10\,170\tilde{\lambda}_0 - 900\tilde{g}_0 - 90\Delta\tilde{A},\\
\langle\sigma^8\rangle &= 105 - 133\,560\tilde{\lambda}_0 - 10\,080\tilde{g}_0 - 840\Delta\tilde{A}.
\end{aligned} \quad (5.16)$$

By Eq. (2.7), $\langle\sigma^2\rangle = 1$, we solve for

$$\Delta\tilde{A} = -6\tilde{g}_1 - 45\tilde{\lambda}_0. \quad (5.17)$$

So for this choice,

$$\begin{aligned}
\langle\sigma^4\rangle &= 3 - 360\tilde{\lambda}_0 - 24\tilde{g}_0,\\
\langle\sigma^6\rangle &= 15 - 6120\tilde{\lambda}_0 - 360\tilde{g}_0,\\
\langle\sigma^8\rangle &= 105 - 95\,760\tilde{\lambda}_0 - 4940\tilde{g}_0,
\end{aligned} \quad (5.18)$$

and thus the cumulants become

$$\begin{aligned}
M_2^0 &= 1,\\
M_4^0 &= -360\tilde{\lambda}_0 - 24\tilde{g}_0,\\
M_6^0 &= -720\tilde{\lambda}_0,\\
M_8^0 &= 0,
\end{aligned} \quad (5.19)$$

plus, of course, terms of order $\tilde{g}_0^2$, $\tilde{g}_0\tilde{\lambda}_0$, and $\tilde{\lambda}_0^2$. To compute (5.10) and (5.15), I need χ and ξ to zero order, and $\partial^2\chi/\partial H^2$ and $\partial^4\chi/\partial H^4$ to first order in $\tilde{g}_0$ and $\tilde{\lambda}_0$. Hence I need to consider only those high-temperature graphs that have any number of vertices that are the meet of two lines, plus one four-line or one six-line vertex. Baker and Kincaid[5] have computed the four-line vertex case. In this counting of lines a derivative $\partial/\partial H_r$ at a point counts as a line, and there are, of course [see (5.12)], four derivatives of $\ln Z$ in $\partial^2\chi/\partial H^2$ and six in $\partial^4\chi/\partial H^4$. The class of graphs with just one six-line vertex to be considered for $\partial^2\chi/\partial H^2$ is just (a) polygons with one root at which four derivatives act, and (b) polygons with one root at which one, two, three, or four linear chains are attached. In case (b) one derivative acts at the free end of each linear chain and the remainder at the root point. We will do the counting on the hyper-simple-cubic lattice. The generating functions are

$$G = 8K/(1-8K),$$

$$P_4(K) = \frac{1}{(2\pi)^d}\int_0^{2\pi}\!\!\int\!\!\int\!\!\int\frac{\Pi_{\tau=1}^4 d\theta_\tau}{[1-(2K\Sigma_{\tau=1}^4\cos\theta_\tau)^2]} - 1, \quad (5.20)$$

for linear chains and polygons in four dimensions, respectively. Thus the sum of this class of graphs is

$$M_6^0 P_4(K)\left[\sum_{j=0}^4\frac{1}{j!}\left(\frac{8K}{1-8K}\right)^j\right]. \quad (5.21)$$

The graphs that contribute to $\partial^4\chi/\partial H^4$ are a single root point with $l = 1,\dots,6$ derivatives and $6-l$ linear chains with a derivative at the free end. The sum of these graphs is

$$M_6^0(1-8K)^{-6}. \quad (5.22)$$

When these results are combined with those of Baker and Kincaid,[5] $\chi = (1-8K)^{-1}$, $\xi^2 = K(1-8K)^{-2}$, we can compute (hyper-simple-cubic lattice), to linear order in $\tilde{g}_0$ and $\tilde{\lambda}_0$,

$$g = K^{-2}\{24\tilde{g}_0 + 360\tilde{\lambda}_0 + 30\tilde{\lambda}_0 P_4(K)[24 - 672K + 7424K^2 - 36\,864K^3 + 69\,632K^4]\}, \quad (5.23)$$

$$\lambda = K^{-4}6!\tilde{\lambda}_0(1-8K). \quad (5.24)$$

By means of (2.6) and (4.2) we note that

$$\lambda = 90\lambda_0/m^2, \quad (5.25)$$

in terms of the parameters of (2.1). As is evident from (5.24), (5.25), and of course (2.6), as $K\to\frac{1}{8}$, which is the continuum limit as $\xi^2\to\infty$ at that value, the value of λ goes to zero in first order. In addition, for the "phantom field" prescription $\tilde{\lambda}_0 = O(1)$ compared to a, $\lambda_0\to 0$ as $a\to 0$

($\xi \to \infty$ with a fixed mass m) and so disappears "like a phantom" in the continuum limit. Needless to say, the expansion of λ does not vanish in all orders in the continuum limit for terms involving $\tilde{\lambda}_0$. An example is given in Fig. 1.

Nevertheless, referring to (5.23) one sees that if I choose, as I am perfectly free to do,

$$\tilde{g}_0 < -(15 + \tfrac{5}{4}P_4(\tfrac{1}{8}))\tilde{\lambda}_0, \tag{5.26}$$

I can make g negative in leading order in perturbation theory for all $0 < K \leqslant K_c = \frac{1}{8}$. That this result is uniformly true in that range of K follows from (5.20) and (5.21) because (i) $P_4(K)/K^2$ is monotonically increasing in K to a finite value at $K = \frac{1}{8}$, and (ii) K^2 times the term in square brackets in (5.21) times $(1 - 8K)^4$ is uniformly bounded by its value at $K = \frac{1}{8}$. This result is in sharp contrast to the case where $\tilde{g}_0 > 0$, where g is positive to leading order in $\tilde{g}_0$ and $\tilde{\lambda}_0$. This result demonstrates, as $g \geqslant 0$ without need for a perturbation expansion by the Lebowitz inequalities[47,48] for $\tilde{\lambda}_0 \equiv 0$, that even though λ_0 is a "phantom" its effects are quite clearly nonvanishing, at least in lowest-order perturbation theory.

Although I have no proof, there are indications that this field theory is not asymptotically free. Namely the contribution of the Feynman diagram shown in Fig. 2 is proportional to

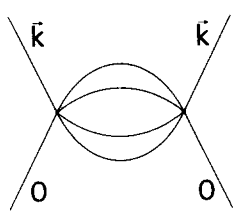


FIG. 2. A phantom field Feynman diagram.

$$(a^2\tilde{\lambda}_0)^2 \int_{-\pi/a}^{\pi/a} \cdots \int d\mathbf{k}_1\, d\mathbf{k}_2\, d\mathbf{k}_3\, \pi(\mathbf{k}_1)\pi(\mathbf{k}_2)\pi(\mathbf{k}_3) \times \pi(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2 - \mathbf{k}_3), \tag{5.27}$$

where the propagator is

$$\pi(\mathbf{p}) = \left[m^2 + \frac{4}{a^2}\sum_{\tau=1}^{4} \sin^2\!\left(\frac{1}{2}\mathbf{p}\cdot\mathbf{e}_\tau a\right)\right]^{-1}, \tag{5.28}$$

and $\mathbf{e}_\tau$ are unit vectors along the crystal axes. This term (5.27) in the continuum limit $a \to 0$ is finite and independent of $\mathbf{k}$. The existence of this $\mathbf{k}$-independent term, and, of course, many others, suggests that the scattering at high momentum does not vanish as in the free-field case and so the theory is not asymptotically free. We remark that this term is numerically different from the corresponding result for the hyper-body-centered cubic lattice and so that theory is probably cutoff dependent.

A further consideration of the momentum dependence of the perturbation expansions suggests that the integration over the propagators can be considered in two cases. First, if the integration over $\mathbf{p}$ converges, then in the continuum limit, $\pi(\mathbf{p}) \to [m^2 + \mathbf{p}^2]^{-1}$ so that rotational invariance holds directly. Second, the integration over $\mathbf{p}$ does not converge in the sense that the integral is made finite only by the existence of limits to the range of integration. For the Feynman diagram expansion of a mass renormalized theory only logarithmically divergent primitive graphs remain. An example is illustrated in Fig. 3. The direct expansion of the propagator to second order in $(\mathbf{p}a)$ leads in this case to a constant that is independent of $\mathbf{p}$ and diverges like $(-\ln a)$ as $a \to 0$ plus a finite part that depends on $\mathbf{p}$ in a rotationally invariant manner in the limit as $a \to 0$. Use was made of the lattice symmetry to derive this result. Elaboration of this result following the by now standard methods of Feynmann diagrammatic expansions lead to the conclusion that the series expansion in $\tilde{\lambda}_0$, while not finite term by term, is at least rotationally invariant. Therefore although, as we have discussed, the direct summation of the $\tilde{\lambda}_0$ expansion in the continuum limit has not been established and so the conclusion of rotational invariance of the theory cannot be deduced from the term-by-term rotational invariance of the $\tilde{\lambda}_0$ series as it could in ϕ_2^4 and ϕ_3^4 theories,[28] nevertheless rotational invariance is not ruled out by the continuum limit of the $\tilde{\lambda}_0$ series. If, in fact, rotational invariance does hold, then all of Nelson's[16] axioms are valid for the field theories we have been studying and, by his reconstruction theorem, correspond to Minkowski space quantum boson field theories satisfying the Wightman axioms.

ACKNOWLEDGMENTS

The author is pleased to acknowledge helpful discussions with F. Cooper, G. Guralnik, J. D. Johnson, J. L. Lebowitz, J. R. Klauder, and A. S. Wightman.

This work was performed under the auspices of the United States Department of Energy.

APPENDIX: MASS GAP

In this Appendix I discuss in more detail the limits involved in the definition of the mass gap. First, to see in Sec. III that the definitions of the physical mass (3.20) in the infinite volume limit agree with the infinite volume limit of the pseudomass (3.21), I argue as follows.

For any box of size L, by (3.21) with fixed lattice spacing (fixed K),

$$\mu(L) \leqslant -\ln[\langle\phi_\mathbf{r}\phi_\mathbf{s}\rangle_L/B]/|\mathbf{r} - \mathbf{s}|, \tag{A1}$$

with equality at the minimum pair $(\mathbf{r},\mathbf{s})$. Take any desired $(\mathbf{r},\mathbf{s})$ fixed and take the limit $L \to \infty$ of (A1). This procedure yields

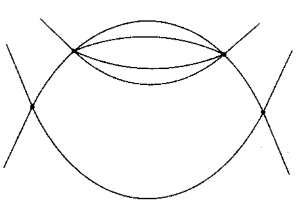


FIG. 1. A six-external leg diagram involving six-line vertices which does not vanish in the continuum limit.

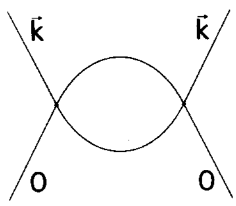


FIG. 3. A primitive, logarithmically divergent diagram.

$$\mu \leqslant -\ln[\langle \phi_{\mathbf{r}}\phi_{\mathbf{s}}\rangle_\infty / B]/|\mathbf{r}-\mathbf{s}|. \tag{A2}$$

Now set $\mathbf{s}=0$ and take the limit,

$$\mu \leqslant \liminf_{\mathbf{r}\to\infty} \frac{-\ln[\langle \phi_0\phi_{\mathbf{r}}\rangle_\infty / B]}{|\mathbf{r}|} = m, \tag{A3}$$

which establishes an inequality between μ and m.

Next I remark that the limit (3.23) selects a sequence $(\mathbf{r}_L, \mathbf{s}_L)$ for which $|\mathbf{r}_L - \mathbf{s}_L| \to \infty$, as proved in Sec. III. If we define

$$\hat{\mu}(L) = -\ln[\langle \phi_{\mathbf{r}_L}\phi_{\mathbf{s}_L}\rangle_\infty / B]/|\mathbf{r}_L - \mathbf{s}_L|, \tag{A4}$$

then, if I select Dirichlet boundary conditions, it has been shown[19] that the two-point function is monotonically increasing in L so $\mu(L) \geqslant \hat{\mu}(L)$. Thus,

$$\mu = \lim_{L\to\infty} \mu(L) \geqslant \lim_{L\to\infty} \hat{\mu}(L) = \hat{\mu}. \tag{A5}$$

However, by translational invariance of the $\langle\ \rangle_\infty$, the limit in (A5) is just a particular sequence in the limit (3.20) and so necessarily $\hat{\mu} \geqslant m$. Thus $\mu \geqslant m$ and hence, by (A3), $\mu = m$, which completes the proof of their equality.

In Sec. IV a more involved limiting process is used. The argument establishing (A3) is similar to that given above. Now B is chosen as a uniform upper bound over the closure $\mathscr{K}$ of the set of K_j. To make the discussion clearer we denote the lattice points by $\mathbf{i}_j$ and $\mathbf{k}_k$, where $\mathbf{r} = a_j \mathbf{i}_j$ and $\mathbf{s} = a_j \mathbf{k}_j$. Equation (A1) becomes

$$\mu(L_j, K_j) \leqslant \frac{-\ln[\langle \phi_{a_j \mathbf{i}_j}\phi_{a_j \mathbf{k}_j}\rangle_{L_j K_j} / B]}{a_j|\mathbf{i}_j - \mathbf{k}_j|}, \tag{A6}$$

again with equality holding for the minimum pair $(\mathbf{i}_j, \mathbf{k}_j)$. In this case we fix $\mathbf{r}$ and $\mathbf{s}$ (let $\mathbf{i}$, $\mathbf{k}$ vary inversely with a_j), and take the limit $j \to \infty$. Since (A6) holds uniformly in j, this limit yields

$$\mu = \lim_{j\to\infty} \mu(L_j, K_j) \leqslant \frac{-\ln[\langle \phi_{\mathbf{r}}\phi_{\mathbf{s}}\rangle_{j=\infty} / B]}{|\mathbf{r}-\mathbf{s}|}. \tag{A7}$$

As for (A3), it follows that

$$\mu \leqslant \liminf_{\mathbf{r}\to\infty} \frac{-\ln[\langle \phi_0\phi_{\mathbf{r}}\rangle / B]}{|\mathbf{r}|} = m. \tag{A8}$$

Next, I show that the appropriate Schwinger functions exist in the limit $j\to\infty$. As I have remarked in Sec. III, Schrader[29] has shown that the two-point function is a monotonic decreasing function of each component $|x_i|$ $[\mathbf{r} = (x_1, x_2, x_3, x_4)]$ separately. In addition it follows from Schrader's[29] Corollary 3.6 that if we write the n-point Schwinger function S_n, so that it is the expectation value of the n ϕ's arranged in nondecreasing order of their ith $(=1,\ldots,4)$ position coordinate, then S_n is monotonic decreasing as a function of the differences of all the successive values of that coordinate. Of course, this result holds whatever i was chosen and so for each i.

By the log convexity [(3.28) and (3.29)] and monotonicity of the two-point function $\langle \psi_0\psi_{\mathbf{r}}\rangle_L$, any limit $j\to\infty$ must be continuous except perhaps at $\mathbf{r}=0$. For, if these were a discontinuity for $|\mathbf{r}_0| > 0$ the derivative would be infinite and there would be by log convexity a region $0 < \mathbf{r} < \mathbf{r}_0$, where $\langle \psi_0 v_{\mathbf{r}}\rangle_\infty = \infty$, which would contradict the proved amplitude renormalization. That such a limit exists follows from monotonicity in the following way. Choose a denumerable set of points, dense in $\{R^4 \setminus |\mathbf{r}| < \epsilon\}$. From (3.45), Griffiths inequality, and monotonicity, for any point $\mathbf{r}$, $|\mathbf{r}| \geqslant \epsilon > 0$,

$$0 \leqslant \langle \psi_0\psi_{\mathbf{r}}\rangle \leqslant 4/(\pi \hat{m}\epsilon^2)^2. \tag{A9}$$

Thus, at the first point there must exist a subsequence of the j's that converge to a limit by standard arguments. From this subsequence we can choose one that converges at the second point, and so on. Thus there exists a subsequence that converges on a dense set and by the argued continuity everywhere in $\{P^4 \setminus |\mathbf{r}| < \epsilon\}$. But as ϵ is arbitrary we can define $\langle \psi_0\psi_{\mathbf{r}}\rangle$ in R^4 punctured at $\mathbf{r}=0$. By two-point dominance we can bound all the higher-order Schwinger functions (on disjoint points) and by a repeat of the above arguments, using the above-quoted monotonicity, we establish the convergence for all the higher-order Schwinger functions on a denumerable, dense set of points. This result suffices to prove convergence at least in the sense of distributions. The final subsequence obtained is what we now choose to be the original sequence of (4.6).

To obtain control of m from above, in addition to the lower bound (A8), I use the results of Simon.[3] The main hypotheses of his theorems are the theory of Markov fields and the FKG inequalities, which are also available for the present case. The point is to study the spectral properties of the transfer matrix $\mathsf{T} = e^{-\mathsf{H}}$. The idea of the proof is that the eigenvalue of the first excited state of H decreases continuously towards the smallest eigenvalue zero, as K increases for finite L and nonzero lattice spacing. Thus one always can select that eigenvalue to correspond to any desired mass gap. The decay of the two-point function is expected to follow closely this first excited state eigenvalue.

The eigenvalue of the transfer matrix is related to the value of the two-point function along the direction perpendicular to the lattice hyperplane defining the transfer matrix. In the notation of (3.24), we define

$$\mu_\perp(L) = \min_{\mathbf{r},n} (-\ln[\langle \phi_{\mathbf{r}}\phi_{\mathbf{r}+n\boldsymbol{\delta}'}\rangle / B]/n|\boldsymbol{\delta}'|). \tag{A10}$$

Monotonicity tells us that other correlation functions $\langle \phi_{\mathbf{r}}\phi_{\mathbf{s}}\rangle$ with the same projected difference of $(\mathbf{s}-\mathbf{r})$ in direction $\boldsymbol{\delta}'$ are less than or equal to $\langle \phi_{\mathbf{r}}\phi_{\mathbf{r}+n\boldsymbol{\delta}'}\rangle$. If we combine this result with lattice symmetry, we find the worst case is on the diagonal, so

$$\mu_\perp(L) \geqslant \mu(L) \geqslant \mu_\perp(L)/\sqrt{d} = \tfrac{1}{2}\mu_\perp(L), \tag{A11}$$

where if rotational invariance holds, as I think likely but have not proved, $\mu(L) = \mu_\perp(L)$. The first inequality in (A11) follows by comparison of (3.21) with (A10).

In the work of Simon[3] rotational invariance is not used in any essential way as he works in a box finite in spacelike directions (1 in his case but he points out his results extend to higher dimensions) and infinite in the imaginary-timelike direction ($\boldsymbol{\delta}'$ here). Simon proves that the first excited state is coupled to the single field operator ϕ. If Ω is the lowest eigenstate (vacuum) of H [his Theorem 6 establishes its existence for our case by (A8)] of eigenvalue 0 [see (3.25)], then he defines $\psi_\perp = \psi - (\Omega, \psi)\Omega$. The unique spectral measure for H associated with $\psi_\perp$ is denoted by $d\mu_{\psi_\perp}$. He further defines $M(\phi) \equiv \inf(\operatorname{Supp} d\mu_{\psi_\perp})$. Then if

$E \equiv \inf[\sigma(H \restriction \{\Omega\}^{\perp}))$ and S is a subset of the Hilbert space $\mathscr{H}$ over which H is defined, when $\inf\{M(\psi)|\psi \in S\} = E$, Simon says S is coupled to the first excited state. Under the conditions $\mu(L) = \hat{m}$, we have, by use of (A8), that Simon's results hold uniformly for every j in (4.6). By log convexity,

$$\mu_{\perp}(L_j) \geqslant E(L_j) > 0, \tag{A12}$$

as $E(L_j)$ is defined by the limit $n \to \infty$ in (A10) and for $\mu_{\perp}(L_j)$ n is finite.

Direct computation of the behavior of any eigenvalues $T_{\alpha}(L_j)$ of the transfer matrix T by standard perturbation theory, shows (i) $\partial T_{\alpha}(L_j)/\partial K$ is bounded (but not uniformly bounded) for finite L, $0 < K < \infty$, and so it is a continuous function of K, and (ii) $T_{\alpha}(L_j)$ is monotonically increasing in K. These results together with Dirichlet boundary conditions and our selection of a sequence j for which the two-point function converges insure that we may take the limit $j \to \infty$ of (A12) to yield

$$\mu_{\perp} \geqslant E \geqslant 0. \tag{A13}$$

By Simon (Theorem 4, Lemma 2), $E = m_{\perp}$. By an argument analogous to (A11) $m_{\perp} \geqslant m \geqslant \frac{1}{2} m_{\perp}$ so

$$2\mu \geqslant m, \tag{A14}$$

or, if rotational invariance is assumed, $\mu \geqslant m$. Thus combining (A8) with (A14) and $\mu = \hat{m}$ by construction we get

$$2\hat{m} \geqslant m \geqslant \hat{m}, \tag{A15}$$

and $m = \hat{m}$ if rotational invariance holds.

The proof of (A8) easily can be adapted to yield $m_{\perp} \geqslant \mu_{\perp}$, which, when coupled with (A13), gives the sharp result $\mu_{\perp} = m_{\perp}$. One thus can choose $m_{\perp} = \hat{m}_{\perp}$ as desired. Since the main body diagonal is perpendicular to a reflection symmetry lattice hyperplane (HCS and HBCC lattices) all the above arguments can be applied to the diagonal direction by use of a different box shape to establish $m_D = \hat{m}_D$, which can be chosen at pleasure instead if desired. Likewise the same results are true for any direction perpendicular to a lattice reflection symmetry hyperplane.

[1] G. A. Baker, Jr. and J. D. Johnson, J. Phys. A. **18**, L261 (1985).

[2] C. Itzykson and J.-B. Zuber, *Quantum Field Theory* (McGraw–Hill, New York, 1980).

[3] B. Simon, Commun. Math. Phys. **31**, 127 (1973).

[4] C. M. Newman, Phys. Lett. B **83**, 63 (1979).

[5] G. A. Baker, Jr., and J. M. Kincaid, J. Statist. Phys. **24**, 469 (1981).

[6] M. Aisenman, Phys. Rev. Lett. **47**, 1 (1981); Commun. Math. Phys. **86**, 1 (1982).

[7] D. Brydges, J. Frölich, and T. Spencer, Commun. Math. Phys. **83**, 123 (1982).

[8] M. Reed, "Abstract non-linear wave equations," in *Lecture Notes in Mathematics*, Vol. 507, edited by A. Dold and B. Eckmann (Springer, Berlin, 1976).

[9] G. A. Baker, Jr., J. Phys. A **17**, L621 (1984).

[10] K. Symanzik, in *Mathematical Problems in Theoretical Physics, Lecture Notes in Physics*, Vol. 153, edited by R. Schrader, R. Seiler, and D. A. Uhlenbrock (Springer, Berlin, 1982).

[11] D. J. Gross and A. Neveu, Phys. Rev. D **10**, 3235 (1974).

[12] B. Lautrup, Phys. Lett. B **69**, 109 (1977).

[13] G. 't Hooft, in *The Whys of Subnuclear Physics*, edited by A. Zichichi (Plenum, New York, 1979).

[14] G. Parisi, Phys. Lett. B **76**, 65 (1978); Phys. Rep. **49**, 215 (1979).

[15] M. C. Bergère and F. David, "Ambiguities of renormalized ϕ_4^4 field theory and the singularities of its Borel transform," Princeton Institute for Advanced Studies preprint, 1983.

[16] E. Nelson, J. Funct. Anal. **12**, 97 (1973).

[17] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and all That* (Benjamin, Reading, MA, 1980).

[18] G. A. Baker, Jr. and J. D. Johnson (submitted for publication).

[19] G. A. Baker, Jr., *Phase Transitions and Critical Phenomena*, edited by C. Domb and J. L. Lebowitz (Academic, London, 1984), Vol. 9, p. 234.

[20] G. A. Baker, Jr., *Essentials of Padé Approximants* (Academic, New York, 1975).

[21] G. A. Baker, Jr. and P. R. Graves-Morris, *Padé Approximants, Part I: Basic Theory*, and *Part II: Extensions and Applications*, in *Encyclopedia of Mathematics and its Applications*, edited by G. C. Rota (Addison–Wesley, Reading, MA, 1981), Vols. 13 and 14.

[22] R. B. Griffiths, J. Math. Phys. **8**, 478, 484 (1967).

[23] J. Ginibre, Commun. Math. Phys. **16**, 310 (1970).

[24] E. Nelson, in *Constructive Quantum Field Theory*, edited by C. DeWitt and R. Stora (Gordon and Breach, New York, 1971).

[25] K. Osterwalder and R. Schrader, Helv. Phys. Acta. **46**, 277 (1973).

[26] K. Osterwalder and R. Schrader, Comm. Math. Phys. **31**, 83 (1973).

[27] R. Schrader and D. Uhlenbrock, J. Funct. Anal. **18**, 369 (1975).

[28] J. Glimm and A. Jaffe, *Quantum Physics* (Springer, New York, 1981).

[29] R. Schrader, Phys. Rev. B **15**, 2798 (1977).

[30] J. Frölich and E. H. Lieb, Commun. Math. Phys. **60**, 233 (1978).

[31] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre, Commun. Math. Phys. **22**, 89 (1971).

[32] J. Bricmont, J.-R. Fontaine, J. L. Lebowitz, and T. Spencer, Commun. Math. Phys. **78**, 363 (1981).

[33] D. Ruelle, Commun. Math. Phys. **50**, 189 (1976).

[34] J. L. Lebowitz and E. Presutti, Commun. Math. Phys. **50**, 195 (1976); **78**, 151 (1980).

[35] A. D. Sokal, J. Statist. Phys. **25**, 25 (1981).

[36] M. E. Fisher, Phys. Rev. **162**, 480 (1967).

[37] E. W. Montroll and G. H. Weiss, J. Math. Phys. **6**, 167 (1965).

[38] Reference 28, pp. 333 and 334.

[39] R. Peierls, Proc. Cambridge Philos. Soc. **32**, 477 (1936).

[40] R. B. Griffiths, Phys. Rev. **136**, A437 (1964).

[41] R. L. Dobruskin, Sov. Phys.-Dokl. **10**, 111 (1965).

[42] H. van Beijeren and G. S. Sylvester, J. Funct. Anal. **28**, 145 (1978).

[43] S. Graffi, V. Grechi, and B. Simon, Phys. Lett. B **32**, 631 (1970).

[44] Reference 20, Chap. 20.

[45] See, for example, G. E. Uhlenbeck and G. W. Ford, *Studies in Statistical Mechanics*, edited by J. de Boer and G. E. Uhlenbeck (North-Holland, Amsterdam, 1962), Vol. 1, p. 119.

[46] M. Wortis, *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, London, 1974), Vol. 3, p. 113.

[47] J. L. Lebowitz, Commun. Math. Phys. **35**, 87 (1974).

[48] G. A. Baker, Jr., J. Math. Phys. **16**, 1324 (1975).

Formal power series solutions of supersymmetric ($N = 3$) Yang–Mills equations

J. Harnad[a)]
Départment de Mathématiques Appliquées, Ecole Polytechnique de Montréal, C. P. 6079, Succursale "A," Montreal, Québec, Canada H3C 3A7

M. Jacques
Institut de Physique Théorique, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

(Received 13 March 1986; accepted for publication 30 April 1986)

Formal power series solutions of the linear system with one spectral parameter associated to the constraint equations for Yang–Mills superconnections on N-extended super-Minkowski space are considered. For $N = 3$ the integrability equations reduce to the supersymmetric field equations. The method of approach is identical to that used by Takasaki for the self-dual equations, based upon formal power series in a spectral parameter and in spatial variables. The problem is reduced to a linear system of equations for a superfield with values in an ∞-dimensional Grassmann manifold. The formal solution is expressed in terms of data on a $(3|2N)$-dimensional superhypersurface. However, a difficulty arises with respect to the Cauchy problem, which becomes formally solvable only for an extended system, breaking the relativistic invariance through introduction of additional superfields.

I. INTRODUCTION

The maximally extended ($N = 3$ or 4) supersymmetric Yang–Mills theory is remarkable for a number of reasons. At the quantum level, it is an ultraviolet finite theory.[1,2] It may be derived by dimensional reduction from a ten-dimensional theory which is the low energy limit of the open superstring.[3,4] Classically, it may, very similarly to the self-dual Yang–Mills theory, be formulated in a geometrical way that permits identification of the field equations as integrability conditions for a linear, overdetermined system of superfield equations with a complex spectral parameter.[5–8] This is based on the supertwistor correspondence of Witten[5] and Manin.[6]

It is possible that this latter interpretation may eventually lead to results on classical solutions analogous to the instanton[9] and monopole[10] constructions. However, there is already a striking structural similarity to the integrable systems of Zakharov–Shabat type which have been so successfully analyzed through inverse spectral methods.[11,12] In particular, the possibility of determining solutions through the matrix Riemann–Hilbert problem has been noted by several authors.[13–15] For supersymmetric two-dimensional models, a straightforward generalization of soliton methods in the superfield formulation has been shown to yield analogous classes of explicit solutions.[16] Such methods may also prove applicable to the four-dimensional, $N = 3$ extended supersymmetric Yang–Mills theory,[17] giving rise to interesting classes of explicitly determined solutions.

Before beginning such a program, it is useful to carry out some relatively simple formal computations, generalized from the two-dimensional framework. Recently, Takasaki[18] developed a new formulation of the self-dual Yang–Mills equations, characterizing the formal power series solutions in terms of certain infinite-dimensional matrix functions. The method derives from Sato's approach to integrable systems based on flows in infinite-dimensional Grassmann manifolds.[19] It is our purpose to show how this analysis extends to the case of the supersymmetric Yang–Mills equations. We find that for this system the equivalence between linear flows which are well defined with respect to Cauchy data and the associated linear system with one spectral parameter fails unless a supplementary linear equation is added to those governing the flows in the Grassmannian. Unfortunately, this additional condition cannot readily be interpreted in terms of the Cauchy data. If it is omitted, however, the system becomes equivalent to an extended superfield system involving additional nonrelativistic terms that has been introduced by Aref'eva and Volovich,[15] and the formal solutions of this modified system are fully characterized by the linearized flows.

II. CONSTRAINT EQUATIONS AND THE LINEAR SYSTEM

In the following, we shall consider complex, affine, N-extended super-Minkowski space $\widetilde{M}$, which may be identified as $\mathbb{C}^{(4|4N)}$ with coordinates

$$(x^{\alpha\dot\beta},\theta^{\alpha}_{s},\theta^{\dot\beta t})\equiv(x,\theta), \tag{2.1}$$

where $\alpha,\beta,\dot\alpha,\dot\beta\in\{1,2\}$ are spinor indices, $s,t\in\{1,...,N\}$ are internal symmetry indices, $x^{\alpha\dot\beta} = x^{\mu}\sigma_{\mu}{}^{\alpha\dot\beta}$, $\mu = 0,...,3$, expresses the Cartesian coordinates $\{x^{\mu}\}$ in a spinor basis through the Pauli matrices, and $\{\theta^{\alpha}_{s},\theta^{\dot\beta t}\}$ are the usual anticommuting spinorial coordinates. The gauge group will also be complex and may, without loss of generality, be regarded as a subgroup of $\mathrm{Gl}(n,\mathbb{C})$ embedded through some faithful representation. Reduction and reality conditions will not be investigated here. The fermionic right translation vector fields are given by

$$D^{s}_{\alpha}\equiv\frac{\partial}{\partial\theta^{\alpha}_{s}}+i\theta^{\dot\beta s}\,\partial_{\alpha\dot\beta}, \tag{2.2a}$$

[a)] On leave of absence from Department of Pure and Applied Mathematics, Stevens Institute of Technology, Hoboken, New Jersey 07030.

 0022-2488/86/092394-08$02.50

$$D_{\dot\beta t} = -\frac{\partial}{\partial\theta^{\dot\beta t}} - i\theta^{\alpha}_{t}\,\partial_{\alpha\dot\beta}, \tag{2.2b}$$

with

$$\partial_{\alpha\dot\beta} = \frac{\partial}{\partial x^{\alpha\dot\beta}}. \tag{2.2c}$$

The gauge superpotentials $\{A_{\alpha\dot\beta}, A^{s}_{\alpha}, A_{\dot\beta t}\}$ are components of a $gl(n,\mathbb{C})$-valued one-form ω on $\widetilde{M}$, determined in some local gauge from the connection form on the super-Yang–Mills bundle. The covariant derivatives are given by

$$\nabla_{\alpha\dot\beta} \equiv \partial_{\alpha\dot\beta} + A_{\alpha\dot\beta}, \tag{2.3a}$$

$$\nabla^{s}_{\alpha} \equiv D^{s}_{\alpha} + A^{s}_{\alpha}, \tag{2.3b}$$

$$\nabla_{\dot\beta t} \equiv D_{\dot\beta t} + A_{\dot\beta t}, \tag{2.3c}$$

where $A_{\alpha\dot\beta} = \omega(\partial_{\alpha\dot\beta})$, $A^{s}_{\alpha} = \omega(D^{s}_{\alpha})$, and $A_{\dot\beta t} = \omega(D_{\dot\beta t})$. The curvature components $\{F^{st}_{\alpha\beta}, F_{\dot\alpha s\dot\beta t}, F^{s}_{\alpha\dot\beta t}, F^{s}_{\mu\alpha}, F_{\mu\dot\alpha t}, F_{\mu\nu}\}$ are introduced in the usual way:

$$F(X,Y) = [\nabla_X, \nabla_Y]_{\pm} - \nabla_{[X,Y]_{\pm}}, \tag{2.4}$$

where $[\ ,\ ]_{-}$ denotes a commutator and $[\ ,\ ]_{+} = \{\ \}$ an anticommutator. For example,

$$F^{st}_{\alpha\beta} = \{\nabla^{s}_{\alpha}, \nabla^{t}_{\beta}\}, \tag{2.5a}$$

$$F^{s}_{\alpha\dot\beta t} = \{\nabla^{s}_{\alpha}, \nabla_{\dot\beta t}\} + 2i\delta^{s}_{t}\nabla_{\alpha\dot\beta}, \tag{2.5b}$$

etc.

In order to reduce the supersymmetry representations defined by Lie differentiation of superfields with respect to the supersymmetry generators (i.e., left-translation vector fields), one imposes certain constraints that are invariant under supersymmetry and gauge transformations. These correspond to the vanishing of the supercurvature components along super null lines[5,6]:

$$F^{st}_{\alpha\beta} + F^{st}_{\beta\alpha} = 0, \tag{2.6a}$$

$$F_{\dot\alpha s\dot\beta t} + F_{\dot\beta s\dot\alpha t} = 0, \tag{2.6b}$$

$$F^{s}_{\alpha\dot\beta t} = 0. \tag{2.6c}$$

Depending on the value of N, these equations may or may not have dynamical content. For the cases $N = 1$ or 2, they merely determine the higher superconnection and supercurvature components in terms of the leading ones without implying any field equations. For $N = 3$, however, they not only imply all the field equations but in fact are equivalent to them, allowing a unique reconstruction of the constrained superconnection in terms of any given solution.[8] Conversely, all solutions of the field equations may be uniquely determined from solutions to the constraint equations. The precise one–one passage between these sets of data is given in Ref. 8. For the present, we are concerned with characterizing solutions to (2.6a)–(2.6c).

To this end, following Ref. 14, note that (2.6a)–(2.6c) imply the local existence of two $Gl(n,\mathbb{C})$-valued superfields (g,h) such that

$$\begin{aligned} &A^{s}_{1} = g^{-1}D^{s}_{1}\, g, \quad A_{\dot 1t} = g^{-1}D_{\dot 1t}\, g,\\ &A^{s}_{2} = h^{-1}D^{s}_{2}h, \quad A_{\dot 2t} = h^{-1}D_{\dot 2t}h,\\ &A_{1\dot 1} = g^{-1}\partial_{1\dot 1}\, g, \quad A_{2\dot 2} = h^{-1}\partial_{2\dot 2}h. \end{aligned} \tag{2.7}$$

After a gauge transformation by g^{-1}, the transformed potentials vanish,

$$A_{1\dot 1} = 0, \quad A^{s}_{1} = 0, \quad A_{\dot 1t} = 0, \tag{2.8}$$

and, in terms of the gauge invariant quantity

$$B = gh^{-1}, \tag{2.9}$$

we have

$$A_{2\dot 2} = B\,\partial_{2\dot 2}B^{-1}, \tag{2.10a}$$

$$A^{s}_{2} = BD^{s}_{2}B^{-1}, \tag{2.10b}$$

$$A_{\dot 2t} = BD_{\dot 2t}B^{-1}. \tag{2.10c}$$

The remaining equations in (2.6) then read

$$D^{s}_{1}(BD^{t}_{2}B^{-1}) + D^{t}_{1}(BD^{s}_{2}B^{-1}) = 0, \tag{2.11a}$$

$$D_{\dot 1s}(BD_{\dot 2t}B^{-1}) + D_{\dot 1t}(BD_{\dot 2s}B^{-1}) = 0, \tag{2.11b}$$

$$D^{s}_{1}(BD_{\dot 2t}B^{-1}) + 2i\delta^{s}_{t}A_{1\dot 2} = 0, \tag{2.11c}$$

$$D_{\dot 1t}(BD^{s}_{2}B^{-1}) + 2i\delta^{s}_{t}A_{2\dot 1} = 0. \tag{2.11d}$$

In this gauge, the constraint equations appear as the integrability conditions for the following linear superfield system:

$$(\lambda D^{s}_{1} + D^{s}_{2} + A^{s}_{2})R = 0, \tag{2.12a}$$

$$(\lambda^{2}D_{\dot 1t} + D_{\dot 2t} + A_{\dot 2t})R = 0, \tag{2.12b}$$

$$\begin{aligned}(\lambda^{3}\partial_{1\dot 1} &+ \lambda^{2}\partial_{2\dot 1} + \lambda\,\partial_{1\dot 2} + \partial_{2\dot 2}\\ &+ \lambda^{2}A_{2\dot 1} + \lambda A_{1\dot 2} + A_{2\dot 2})R = 0,\end{aligned} \tag{2.12c}$$

where R is an invertible matrix function on $\widetilde{M}$ depending on the spectral parameter λ.

The system (2.12) is formulated this way in Refs. 7 and 14. It may be interpreted as covariant constancy conditions conditions for sections along super null lines, restricted to a certain $(5|4N)$-dimensional submanifold within the $(6|4N)$-dimensional supermanifold of pairs of $\{$points $p\in\widetilde{M}$, super null lines through $p\}$. However, it would appear that no relevant data are lost by this restriction, since the integrability conditions are the same as on the full $(6|4N)$ space, and the superconnection satisfying (2.10) and (2.11) can be uniquely determined from R. In the next section, we shall derive these solutions by considering formal power series expansions of R in the spectral parameter.

III. SPECTRAL PARAMETER EXPANSIONS AND THE *M*-MATRIX

We now follow a procedure analogous to that used by Takasaki[18] for obtaining formal power series solutions of the self-dual Yang–Mills equations. Since much of the analysis is logically equivalent to Ref. 18, we shall omit many of the details. Consider solutions to (2.12) in the form of a formal power series in λ^{-1}:

$$R(\lambda) = \sum_{j=0}^{\infty} R_j\,\lambda^{-j}, \tag{3.1}$$

with normalization

$$R_0 = \mathbf{1}. \tag{3.2}$$

[To make precise many of the procedures implicit in the following, the coefficients R_j must also be regarded as formal series in the coordinates (x,θ).]

Introducing (3.1) in (2.12) and identifying like powers of λ gives

$$D^{s}_{1}\,R_{j+1} + D^{s}_{2}\,R_j + A^{s}_{2}\,R_j = 0, \tag{3.3a}$$

$$D_{\dot 1t}\,R_{j+2} + D_{\dot 2t}\,R_j + A_{\dot 2t}\,R_j = 0, \tag{3.3b}$$

$$\partial_{1\dot{1}} R_{j+3} + \partial_{2\dot{1}} R_{j+2} + \partial_{1\dot{2}} R_{j+1} + \partial_{2\dot{2}} R_j + A_{2\dot{1}} R_{j+2} + A_{1\dot{2}} R_{j+1} + A_{2\dot{2}} R_j = 0. \tag{3.3c}$$

These equations are also valid for $j<0$, provided we set $R_j = 0$ for $j<0$. The $j=0$, -1, and -2 equations allow us to express the connection components in terms of the leading coefficients of the expansion as

$$A_2^s = -D_1^s R_1, \tag{3.4a}$$

$$A_{\dot{2}t} = -D_{\dot{1}t} R_2, \tag{3.4b}$$

$$A_{2\dot{1}} = -\partial_{1\dot{1}} R_1, \tag{3.4c}$$

$$A_{1\dot{2}} = -\partial_{1\dot{1}} R_2 - \partial_{2\dot{1}} R_1 - A_{2\dot{1}} R_1, \tag{3.4d}$$

$$A_{2\dot{2}} = -\partial_{1\dot{1}} R_3 - \partial_{2\dot{1}} R_2 - \partial_{1\dot{2}} R_1 - A_{2\dot{1}} R_2 - A_{1\dot{2}} R_1 \tag{3.4e}$$

and also imply the equation

$$D_{\dot{1}t} R_1 = 0. \tag{3.4f}$$

Thus (3.3a)–(3.3c), with the connection components given by (3.4a)–(3.4e), may be regarded as a nonlinear superfield system for R, whose solution yields the solutions of (2.8)–(2.11) determined by (3.4a)–(3.4e).

To solve the system (3.3a)–(3.3c), it is convenient also to introduce the formal expansion for $R^{-1}(\lambda)$,

$$R^{-1}(\lambda) = \sum_{j=0}^{\infty} R_j^* \lambda^{-j}, \tag{3.5}$$

where $R_0^* = 1$ and the remaining coefficients $\{R_j^*\}$ are uniquely recursively determined from those of the expansion (3.1) by the condition

$$R(\lambda)R^{-1}(\lambda) = R^{-1}(\lambda)R(\lambda) = 1. \tag{3.6}$$

Instead of considering either the system (3.3a)–(3.3c) or the correpsonding system for $\{R_j^*\}$, we introduce, as in Ref. 18, an ∞-dimensional matrix M, consisting of $n\times n$ dimensional blocks $\{M_{i,-j-1}\}_{i=0,\ldots,\infty,j=0,\ldots,\infty}$ defined by the generating function expansion

$$R^{-1}(\mu)R(\lambda) \equiv 1 + (\lambda-\mu) \sum_{i,j=0}^{\infty} M_{i,-j-1}\mu^{-i-1}\lambda^{-j-1}. \tag{3.7}$$

(This definition should be compared with the finite-dimensional M-matrix defined for the soliton sector of integrable systems of Zakharov–Shabat type in Refs. 16 and 20.) More explicitly, $M_{i,-j-1}$ is defined by the bilinear sums:

$$M_{i,-j-1} = \sum_{k=-j-1}^{-1} R^*_{i-k} R_{k+j+1}. \tag{3.8a}$$

This suggests extending the definition to negative i as

$$M_{ij} = \delta_{ij}, \quad i,j<0. \tag{3.8b}$$

It follows, by equating leading terms in (3.7), that $R(\lambda)$ and $R^{-1}(\mu)$ are determined from M_{ij} by

$$R(\lambda) = 1 - \sum_{j=0}^{\infty} M_{0,-j-1}\lambda^{-j-1}, \tag{3.9a}$$

$$R^{-1}(\mu) = 1 + \sum_{i=0}^{\infty} M_{i,-1}\mu^{-i-1}, \tag{3.9b}$$

and, consequently, that M_{ij} satisfies the quadratic constraints

$$M_{i+1,-j-1} - M_{i,-j-2} = M_{i,-1} M_{0,-j-1}. \tag{3.10}$$

Conversely, given a semi-infinite matrix M satisfying the constraints (3.10), it follows that it may be determined from its 0th row and (-1)th column through relations (3.9a), (3.9b), (3.7), and (3.8). The fact that $R(\lambda)$ and $R^{-1}(\lambda)$ so defined really are formal inverses of each other also follows from Eqs. (3.7) and (3.8).

We can now derive a set of equations for M that are equivalent to (3.3), whose solution therefore determines the superconnection components satisfying (2.10) and (2.11). Applying the operators $\mu D_1^s + D_2^s$, $\mu^2 D_{\dot{1}t} + D_{\dot{2}t}$, and $\mu^3\partial_{1\dot{1}} + \mu^2\partial_{2\dot{1}} + \mu\,\partial_{1\dot{2}} + \partial_{2\dot{2}}$ to both sides of Eq. (3.7), and using Eqs. (2.12), (3.9), and (3.10), we deduce that M satisfies the equations

$$D_1^s M_{i+1,-j-1} + D_2^s M_{i,-j-1} - M_{i,-1} D_1^s M_{0,-j-1} = 0, \tag{3.11a}$$

$$D_{\dot{1}t} M_{i+2,-j-1} + D_{\dot{2}t} M_{i,-j-1} - M_{i,-1} D_{\dot{1}t} M_{1,-j-1} - M_{i,-2} D_{\dot{1}t} M_{0,-j-2} = 0, \tag{3.11b}$$

$$\begin{aligned}&\partial_{1\dot{1}} M_{i+3,-j-1} + \partial_{2\dot{1}} M_{i+2,-j-1} + \partial_{1\dot{2}} M_{i+1,-j-1} \\ &+ \partial_{2\dot{2}} M_{i,-j-1} - M_{i+2,-1}\partial_{1\dot{1}} M_{0,-j-1} \\ &- M_{i+1,-1}(\partial_{1\dot{1}} M_{0,-j-2} + \nabla_{2\dot{1}} M_{0,-j-1}) \\ &- M_{i,-1}(\partial_{1\dot{1}} M_{0,-j-3} + \nabla_{2\dot{1}} M_{0,-j-2} + \nabla_{1\dot{2}} M_{0,-j-1}) \\ &= 0,\end{aligned} \tag{3.11c}$$

plus the additional equation

$$D_{\dot{1}t} M_{0,-1} = 0, \tag{3.11d}$$

which follows from (3.4f). For $i=0$, in view of Eq. (3.9a) and the constraints (3.10), Eq. (3.11) reproduces (3.3a)–(3.3c) and (3.4a)–(3.4f). We thus have:

Proposition 3.1: The system of equations (3.3a)–(3.3c) and (3.4a)–(3.4f) is equivalent to the system (3.11)–(3.11d), with $M_{i,-j-1}$ subject to the constraints (3.10).

It may appear as if little is gained by this, since the system (3.11a)–(3.11c) seems, if anything, more difficult to analyze than (3.3a)–(3.3c). We shall see, however, in the following section that the matrix $M_{i,-j-1}$, together with Eqs. (3.11a)–(3.11c) and constraints (3.10) have a simple interpretation in terms of infinite-dimensional Grassmann manifolds, which leads to a linearization and formal integration of the system.

IV. GRASSMANN MANIFOLDS AND LINEARIZATION

We now turn to a geometrical interpretation of the M-matrix which will be used to linearize and formally integrate Eqs. (3.11a)–(3.11c). Consider the linear space $\mathbb{C}[[\lambda,\lambda^{-1}]]\otimes\mathbb{C}^n \equiv V$ of formal Laurent series in the variables (λ,λ^{-1}) with values in $\mathbb{C}^n$; i.e.,

$$V = \left\{ \sum_{i=-l}^{+\infty} v_i\lambda^i, \quad v_i\in\mathbb{C}^n, \quad l\in\mathbb{Z} \right\}.$$

Decomposing V into components consisting of positive and negative powers

$$V = V_+ + V_-,$$

$$V_- \equiv \left\{ \sum_{i=-l}^{-1} v_i\lambda^i \in V \right\}, \quad V_+ \equiv \left\{ \sum_{i=0}^{+\infty} v_i\lambda^i \in V \right\},$$

we may correspondingly decompose the endomorphisms

$$\mathrm{End}(V) = \mathrm{End}(V_+) + \mathrm{Hom}(V_+,V_-) + \mathrm{Hom}(V_-,V_+) + \mathrm{End}(V_-).$$

We shall refer to the groups of invertible endomorphisms of V, V_+, and V_-, respectively, as $\mathrm{Gl}(2n\infty)$, $\mathrm{Gl}^+(n\infty)$, and $\mathrm{Gl}^-(n\infty)$. The Lie algebra $\mathrm{gl}(2n\infty)$ may be identified as $\mathrm{End}(V)$, with the Lie bracket defined by commutators of maps. Referring all maps to the natural basis $\{\lambda^i \otimes e_j\}_{i\in\mathbb{Z}, j=1,\ldots,n}$, where $\{e_j\}$ is the standard $\mathbb{C}^n$ basis, we have the matrix representation

$$\mathrm{End}(V) \sim \left\{ \begin{array}{c|cc} & -\infty \quad -1 & 0 \quad \infty \\ \hline \begin{matrix} -\infty \\ -1 \\ 0 \\ \infty \end{matrix} & \multicolumn{2}{c}{\left(\begin{array}{c|c} T_{--} & T_{-+} \\ \hline T_{+-} & T_{++} \end{array} \right)} \end{array} \right\},$$

where the ranges of indices are as indicated:

$$\begin{aligned} T_{--} &= \{T_{ij}, \quad i,j<0\} && \sim \mathrm{End}(V_-), \\ T_{-+} &= \{T_{ij}, \quad i<0, \quad j\geqslant 0\} && \sim \mathrm{Hom}(V_+,V_-) \\ T_{+-} &= \{T_{ij}, \quad i\geqslant 0, \quad j<0\} && \sim \mathrm{Hom}(V_-,V_+), \\ T_{++} &= \{T_{ij}, \quad i,j\geqslant 0\} && \sim \mathrm{End}(V_+). \end{aligned}$$

To make sense as maps on V, each column in T_{--} and each row in T_{-+}, T_{++} must have only finitely many nonvanishing elements, and all but a finite number of rows in T_{-+} must vanish. Now, consider the Grassmannian $\mathrm{Gr}_-(2n\infty)$ of subspaces of V modeled on V_-; i.e., the $\mathrm{Gl}(2n\infty)$ orbit of V_- or, equivalently, the images of injective maps $T\in\mathrm{Hom}(V_-,V)$. (A more rigorous formulation of these loop spaces and Grassmannians may be found in Ref. 21.) As in the finite-dimensional case, these may be represented by homogeneous coordinates consisting here of semi-infinite rectangular matrices of the type $\binom{T_{--}}{T_{+-}}$, with linearly independent columns $\{T(\lambda^i)\}_{i=-1,\ldots,-\infty}$ defining a frame for the given space. The points of $\mathrm{Gr}_-(2n\infty)$ are identified with equivalence classes $[T]$ under change of frame; i.e.,

$$[T] = \{T\circ h, \quad h\in\mathrm{Gl}^-(n\infty), \quad T\in\mathrm{Hom}(V_-,V)\}.$$

The affine part $\mathrm{Gr}_-^A(2n\infty)\subset\mathrm{Gr}_-(2n\infty)$, consisting of those points for which T_{--} is invertible, may be identified with $\mathrm{Hom}(V_-,V_+)$ through the affine coordinates $M_{+-} = \{M_{ij}\}_{i\geqslant 0, j<0}$ defined by

$$M_{+-} \equiv T_{+-}T_{--}^{-1}. \tag{4.1}$$

For such points, $T\sim\binom{T_{--}}{T_{+-}}$ has the equivalent representation $\binom{1}{M_{+-}}\equiv M$, $(1)_{ij}=\delta_{ij}$, $i,j<0$.

Now consider the element of $\mathrm{Gl}(2n\infty)$ defined by multiplication by λ^{-1}:

$$\lambda^{-1}: v\in V \to \lambda^{-1}v. \tag{4.2}$$

This has the matrix representation

$$\Lambda = \left(\begin{array}{c|c} \Lambda_{--} & \Lambda_{-+} \\ \hline \Lambda_{+-} & \Lambda_{++} \end{array} \right), \quad \{\Lambda_{ij} = \delta_{i+1,j}\}, \quad i,j\in\mathbb{Z}, \tag{4.3}$$

where

$$\Lambda_{--} = \begin{array}{c} \\ -\infty \\ \\ \\ \\ -1 \\ \end{array} \!\!\begin{array}{c} \left(\begin{matrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \cdot & & & \vdots \\ \cdot & & \ddots & & 0 \\ \cdot & & & \cdot & 1 \\ 0 & \cdot & \cdot & \cdots & 0 \end{matrix} \right) \\ -\infty \qquad\qquad\qquad -1 \end{array},$$

$$\Lambda_{-+} = \begin{array}{c} -\infty \\ \\ \\ \\ -1 \\ \end{array} \!\!\begin{array}{c} \left(\begin{matrix} 0 & \cdot & \cdots & \cdot & 0 \\ \cdot & \cdot & & & \cdot \\ \vdots & & \ddots & & \vdots \\ 0 & & & \cdot & \cdot \\ 1 & 0 & \cdots & \cdot & 0 \end{matrix} \right) \\ 0 \qquad\qquad\qquad \infty \end{array},$$

$$\Lambda_{++} = \begin{array}{c} 0 \\ \\ \\ \\ \infty \\ \end{array} \!\!\begin{array}{c} \left(\begin{matrix} 0 & 1 & \cdots & \cdot & 0 \\ \cdot & \cdot & & & \cdot \\ \vdots & & \ddots & & \vdots \\ \cdot & & & \cdot & 1 \\ 0 & \cdot & \cdots & \cdot & 0 \end{matrix} \right) \\ 0 \qquad\qquad\qquad \infty \end{array}, \quad \Lambda_{+-} = 0.$$

The action of $\mathrm{Gl}(2n\infty)$ on $\mathrm{Gr}_-^A(2n\infty)$ may be expressed in affine coordinates, as usual, by linear fractional transformations:

$$G \equiv \begin{pmatrix} G_{--} & G_{-+} \\ G_{+-} & G_{++} \end{pmatrix},$$

$$G\colon \quad M_{+-} \mapsto (G_{++}M_{+-} + G_{+-})(G_{-+}M_{+-} + G_{--})^{-1}, \tag{4.5}$$

provided the image is also in $\mathrm{Gr}_-^A(2n\infty)$.

We may now interpret the results of the previous section in terms of flows in $\mathrm{Gr}_-(2n\infty)$. We regard the values of the M-matrix as defining the affine coordinates of a point in $\mathrm{Gr}_-(2n\infty)$ as above. The quadratic constraint (3.10) may be expressed in the above notation as

$$\Lambda_{++}M_{+-} = M_{+-}\Lambda_{--} + M_{+-}\Lambda_{-+}M_{+-}. \tag{4.6}$$

This is equivalent to the relation

$$\Lambda[M] = [M] \tag{4.7}$$

or, explicitly

$$\Lambda M = MC, \tag{4.7'}$$

where

$$C \equiv \Lambda_{--} + \Lambda_{-+}M_{+-},$$

which means that the linear space spanned by the columns of M is invariant under the map (4.2). The set of such fixed points in $\mathrm{Gr}_-(2n\infty)$ under this map will be denoted $\mathrm{Gr}_-^\Lambda(2n\infty)$. The set of $X\in\mathrm{End}(V)$ commuting with the map (4.2) leaves $\mathrm{Gr}_-^\Lambda(2n\infty)$ invariant and forms a subalgebra of $\mathrm{gl}(2n\infty)$, which may be identified with the formal loop algebra

$$\widetilde{\mathrm{gl}(n)} \equiv \mathrm{gl}(n) \otimes \mathbb{C}[[\lambda,\lambda^{-1}]]$$

through the map $\Phi\colon \widetilde{\mathrm{gl}(n)} \to \mathrm{gl}(2n\infty)$ defined by

$$\Phi(X\lambda) = X \otimes \Lambda. \tag{4.8}$$

We may now express Eqs. (3.11a)–(3.11c) in a matrix form which is more readily interpretable on $\mathrm{Gr}_-(2n\infty)$.

Proposition 4.1: The system of Eqs. (3.11a)–(3.11c), together with constraints (3.10) is equivalent to the following matrix system:

$$(\Lambda D_1^s + D_2^s)M = MU^s, \tag{4.9a}$$

$$(\Lambda^2 D_{\dot 1 t} + D_{\dot 2 t})M = MV_t, \tag{4.9b}$$

$$(\Lambda^3 \partial_{1\dot 1} + \Lambda^2 \partial_{2\dot 1} + \Lambda \partial_{1\dot 2} + \partial_{2\dot 2})M = MW, \tag{4.9c}$$

where the matrices U^s, V_t, and W are defined by

$$U^s \equiv \Lambda_{-+} D_1^s M_{+-}, \tag{4.10a}$$

$$V_t \equiv (\Lambda_{--}\Lambda_{-+} + \Lambda_{-+}\Lambda_{++}) D_{\dot 1 t} M_{+-}, \tag{4.10b}$$

$$\begin{aligned} W \equiv (&\Lambda^2_{--}\Lambda_{-+} + \Lambda_{--}\Lambda_{-+}\Lambda_{++} \\ &+ \Lambda_{-+}\Lambda^2_{++})\partial_{1\dot 1} M_{+-} \\ &+ (\Lambda_{--}\Lambda_{-+} + \Lambda_{-+}\Lambda_{++})\partial_{2\dot 1} M_{+-} \\ &+ \Lambda_{-+}\partial_{1\dot 2} M_{+-}, \end{aligned} \tag{4.10c}$$

and M satisfies (4.7′).

The proof of this equivalence is a straightforward computation. The upper (T_{--}) block is in fact an identity, in view of Eqs. (4.10a)–(4.10c), whereas the lower (T_{-+}) block may be written

$$\begin{aligned} &\Lambda_{++} D_1^s M_{+-} + D_2^s M_{+-} = M_{+-} U^s, \\ &\Lambda^2_{++} D_{\dot 1 t} M_{+-} + D_{\dot 2 t} M_{+-} = M_{+-} V_t, \\ &(\Lambda^3_{++}\partial_{1\dot 1} + \Lambda^2_{++}\partial_{2\dot 1} + \Lambda_{++}\partial_{1\dot 2} + \partial_{2\dot 2}) M_{+-} \\ &\quad = M_{+-} W. \end{aligned} \tag{4.11}$$

Expressed in components, (4.11) is seen to be equivalent to (3.11a)–(3.11c) in view of the constraint (3.10) and its derivatives. ■

The explicit form of the matrices U^s, V_t, and W given in (4.10a)–(4.10c) is, in fact, irrelevant if Eqs. (4.9a)–(4.9c) are interpreted in $\mathrm{Gr}_-(2n\infty)$. They are determined by the particular choice of affine coordinates in which M takes the form $\binom{1}{M_{+-}}$. Another choice among the homogeneous coordinates, defined by the change of basis

$$M = \begin{pmatrix} 1 \\ M_{+-} \end{pmatrix} \to \widetilde{M} \equiv MH = \begin{pmatrix} H \\ M_{+-}H \end{pmatrix}, \tag{4.12}$$

where H is the matrix representation of some $h \in \mathrm{Gl}_-(n\infty)$, leads to equations of the same form as (4.9) with U^s, V_t, and W replaced by

$$\widetilde{U}^s \equiv H^{-1}U^s H + H^{-1}C(D_1^s H) + H^{-1}D_2^s H,$$

$$\widetilde{V}_t \equiv H^{-1}V_t H + H^{-1}C^2 D_{\dot 1 t} H + H^{-1}D_{\dot 2 t} H,$$

$$\begin{aligned} \widetilde{W} \equiv\; &H^{-1}WH + H^{-1}C^3(\partial_{1\dot 1}H) + H^{-1}C^2(\partial_{2\dot 1}H) \\ &+ H^{-1}C(\partial_{1\dot 2}H) + H^{-1}(\partial_{2\dot 2}H). \end{aligned}$$

The condition (4.7) implying that $M \in \mathrm{Gr}^\Lambda_-(2n\infty)$ still holds, of course, for $\widetilde{M} \equiv MH$, but the coordinate representation (4.7′) is replaced by

$$\Lambda\widetilde{M} = \widetilde{M}\widetilde{C}, \tag{4.7″}$$

where

$$\widetilde{C} \equiv H^{-1}CH.$$

In fact, Eqs. (4.9a)–(4.9c), correctly interpreted, simply state that the superfield function on super-Minkowski space with values in $\mathrm{Gr}_-(2n\infty)$, whose expression in affine coordinates is given by the values of the M-matrix, is annihilated by the differential operators $\Lambda D_1^s + D_2^s$, $\Lambda^2 D_{\dot 1 t} + D_{\dot 2 t}$, and $\Lambda^3\partial_{1\dot 1} + \Lambda^2\partial_{2\dot 1} + \Lambda\partial_{1\dot 2} + \partial_{2\dot 2}$. We summarize this result as follows.

Proposition 4.2: The system (4.9a)–(4.9c) may, by a transformation of the type (4.12), be reduced to the form

$$(\Lambda D_1^s + D_2^s)\widetilde{M} = 0, \tag{4.13a}$$

$$(\Lambda^2 D_{\dot 1 t} + D_{\dot 2 t})\widetilde{M} = 0, \tag{4.13b}$$

$$(\Lambda^3\partial_{1\dot 1} + \Lambda^2\partial_{2\dot 1} + \Lambda\partial_{1\dot 2} + \partial_{2\dot 2})M = 0, \tag{4.13c}$$

where $\widetilde{M} = \binom{\widetilde{M}_{--}}{\widetilde{M}_{++}}$ with $\widetilde{M}_{--}$ invertible. Conversely, all solutions to (49a)–(4.9c) with M satisfying constraint (4.7′) may be deduced from a solution to (4.13a)–(4.13c) by passing to suitable affine coordinates through

$$M = \widetilde{M}\widetilde{M}^{-1}_{--}. \tag{4.14}$$

Geometrically, this equivalence is immediate from the form of Eqs. (4.9a)–(4.9c), since the right-hand side may be interpreted as an infinitesimal change of homogeneous coordinates for the same point. However, it may also be proved explicitly as follows. Notice that (4.9a)–(4.9c), together with (4.7) and the fact that M represents an immersion, implies that $\{U^s, V_t, W\}$ satisfy the integrability conditions

$$\{CD_1^s + D_2^s + U^s, CD_1^t + D_2^t + U^t\} = 0, \tag{4.15a}$$

$$\{C^2 D_{\dot 1 s} + D_{\dot 2 s} + V_s, C^2 D_{\dot 1 t} + D_{\dot 2 t} + V_t\} = 0, \tag{4.15b}$$

$$\begin{aligned} &\{CD_1^s + D_2^s + U^s, C^2 D_{\dot 1 t} + D_{\dot 2 t} + V_t\} \\ &\quad = 2i\delta_t^s[C^3\partial_{1\dot 1} + C^2\partial_{2\dot 1} + C\partial_{12} + \partial_{2\dot 2}] + 2i\delta_t^s W, \end{aligned} \tag{4.15c}$$

$$\begin{aligned} &[CD_1^s + D_2^s + V^s, C^3\partial_{1\dot 1} \\ &\quad + C^2\partial_{2\dot 1} + C\partial_{1\dot 2} + \partial_{2\dot 2} + W] = 0, \end{aligned} \tag{4.15d}$$

$$\begin{aligned} &[C^2 D_{\dot 1 s} + D_{\dot 2 s} + V_s, C^3\partial_{1\dot 1} \\ &\quad + C^2\partial_{2\dot 1} + C\partial_{1\dot 2} + \partial_{2\dot 2} + W] = 0. \end{aligned} \tag{4.15e}$$

It follows that U^s, V_t, and W may be expressed in the form

$$U^s = -[(CD_1^s + D_2^s)H]H^{-1}, \tag{4.16a}$$

$$V_t = -[(C^2 D_{\dot 1 t} + D_{\dot 2 t})H]H^{-1}, \tag{4.16b}$$

$$W = -[(C^3\partial_{1\dot 1} + C^2\partial_{2\dot 1} + C\partial_{1\dot 2} + \partial_{2\dot 2})H]H^{-1}, \tag{4.16c}$$

for some invertible H. Applying transformation (4.12) with this H, the resulting $\widetilde{M}$ satisfies (4.13a)–(4.13c). ■

We now consider the formal Cauchy problem for Eqs. (4.9a)–(4.9c), with $\{U^s, V_t, W\}$ defined by (4.10a)–(4.10c), and M of the form $\binom{1}{M_{+-}}$ satisfying the constraints (4.7). Let $\mathscr{H}$ be the "hyperplane":

$$\mathscr{H} = \{(x,\theta) \mid x^{2\dot 2} = 0, \quad \theta_s^2 = 0, \quad \theta^{\dot 2 t} = 0\}.$$

Consider two solutions to (4.9a)–(4.9c) M_1, M_2 taking the same values on $\mathscr{H}$

$$M_1|_{\mathscr{H}} = M_2|_{\mathscr{H}} = M^0(x^{1\dot 2}, x^{2\dot 1}, x^{1\dot 1}, \theta_s^1, \theta^{\dot 1 t}).$$

We then also have

$$\frac{\partial}{\partial\theta^1_s}M_1|_{\mathscr{H}}=\frac{\partial}{\partial\theta^1_s}M_2|_{\mathscr{H}},\quad \frac{\partial}{\partial\theta^{\dot 1t}}M_1|_{\mathscr{H}}=\frac{\partial}{\partial\theta^{\dot 1t}}M_2|_{\mathscr{H}},$$

and

$$\partial_{\alpha\dot\beta}M_1|_{\mathscr{H}}=\partial_{\alpha\dot\beta}M_2|_{\mathscr{H}},\quad \text{for } (\alpha\dot\beta)\neq(2\dot2).$$

Moreover, since the definitions (4.10a)–(4.10c) do not contain any derivatives with respect to $(x^{2\dot2},\theta^2_s,\theta^{\dot2t})$, the corresponding quantities (U_1^s,V_{1t},W_1), (U_2^s,V_{2t},W_2) derived from M_1, M_2, respectively, also agree on $\mathscr{H}$:

$$U_1^s|_{\mathscr{H}}=U_2^s|_{\mathscr{H}},\quad V_{t1}|_{\mathscr{H}}=V_{t2}|_{\mathscr{H}},\quad W_1|_{\mathscr{H}}=W_2|_{\mathscr{H}}.$$

It follows from Eqs. (4.9a)–(4.9c) that the other derivatives also agree:

$$\partial_{2\dot2}M_1|_{\mathscr{H}}=\partial_{2\dot2}M_2|_{\mathscr{H}},$$

$$\frac{\partial}{\partial\theta^1_s}M_1|_{\mathscr{H}}=\frac{\partial}{\partial\theta^2_s}M_2|_{\mathscr{H}},\quad \frac{\partial}{\partial\theta^{\dot 1t}}M_1|_{\mathscr{H}}=\frac{\partial}{\partial\theta^{\dot 1t}}M_2|_{\mathscr{H}}.$$

Proceeding inductively, we conclude that all higher derivatives also coincide, and hence that if M_1, M_2 are formal power series solutions in (x,θ), they are equal.

Conversely, let $M^0=\binom{1}{M^0_{+-}}$ be given on $\mathscr{H}$, subject to the constraint (4.7). Introduce the new coordinates $(v_1,v_2,v_3,x^{2\dot2},\theta^1_s,\theta^2_s,\theta^{\dot1t},\theta^{\dot2t})$ defined by

$$v_1\equiv x^{1\dot2}+i\theta^1_t\theta^{\dot2t},\tag{4.17a}$$

$$v_2\equiv x^{1\dot1}-i\theta^1_t\theta^{\dot1t},\tag{4.17b}$$

$$v_3\equiv x^{2\dot1}-i\theta^2_t\theta^{\dot1t},\tag{4.17c}$$

and the matrix differential operators:

$$\mathscr{D}_1\equiv-\Lambda^2\left\{(v_1-2i\theta^1_t\theta^{\dot2t})\frac{\partial}{\partial v_2}+(x^{2\dot2}-i\theta^2_t\theta^{\dot2t})\frac{\partial}{\partial v_3}+\theta^{\dot2t}\frac{\partial}{\partial\theta^{\dot1t}}\right\},\tag{4.18a}$$

$$\mathscr{D}_2\equiv-\Lambda\left\{(x^{2\dot2}+i\theta^2_t\theta^{\dot2t})\frac{\partial}{\partial v_1}+\theta^2_s\frac{\partial}{\partial\theta^1_s}\right\},\tag{4.18b}$$

$$\mathscr{D}_3\equiv\Lambda^2v_1\frac{\partial}{\partial v_2}.\tag{4.18c}$$

Now define, in the formal power series sense (cf. Ref. 18),

$$\widetilde M=\exp\mathscr{D}_1\exp\mathscr{D}_2\exp\mathscr{D}_3M^0(v_1,v_2,v_3\theta^1_s,\theta^{\dot1t}).\tag{4.19}$$

Proposition 4.3: The matrix $\widetilde M$ satisfies Eqs. (4.13a)–(4.13c) and the constraint (4.7). Moreover, decomposing

$$\widetilde M\equiv\begin{pmatrix}\widetilde M_{--}\\ \widetilde M_{+-}\end{pmatrix},$$

the infinite matrix $\widetilde M_{--}$ is invertible, in the formal power series sense.

The proof that $\widetilde M$ so defined satisfies Eqs. (4.13a)–(4.13c) follows by repeated application of the identity

$$[D,e^X]=e^X[D,X],$$

valid for operators X,D, with X even, satisfying $[[D,X],X]=0$. The fact that $\widetilde M_{--}$ is invertible and the existence of $\widetilde C$ such that (4.7′) holds is proved using the same arguments as in Ref. 18. ■

Now, combining these results with Proposition 4.2, we conclude that the formal Cauchy problem for Eqs. (4.9a)–(4.9c) has a unique solution for data $M|_{\mathscr{H}}=M^0$ given on the initial data hypersurface $\mathscr{H}$.

Summarizing, we have the following.

Theorem 4.4: The unique formal power series solution to Eqs. (4.9a)–(4.9c), (4.10a)–(4.10c) for M satisfying constraint (4.7), with initial values on $\mathscr{H}$,

$$M|_{\mathscr{H}}=\begin{pmatrix}1\\ M^0_{+-}\end{pmatrix}\equiv M^0,$$

is given by $M=\widetilde M\widetilde M^{-1}_{--}$, where $\widetilde M$ is determined from M^0 by Eq. (4.13).

Note that in all of the above we have established equivalence with the system of Eqs. (3.11a)–(3.11c) only. The formal solutions of the original linear system (2.12a)–(2.12c) with spectral parameter λ are all expressible in the form given by Theorem 4.4. However, the latter do not necessarily satisfy the additional linear equation (3.11d), and hence, for arbitrary initial values M^0, will not determine solutions to (2.12a)–(2.12c) nor, consequently, the superfield constraint equations (2.6a)–(2.6c). On the other hand, the additional equation (3.11d), though it may be added to the linear system of Proposition 4.2 [or equivalently to the system (4.9a)–(4.9c) of Proposition 4.1], does not define an involutive system of first-order operators and hence does not readily translate into additional conditions determining the Cauchy data.

The actual content of Eqs. (4.9a)–(4.9c), without the extra condition, may be expressed in terms of a modified superfield system introduced by Aref'eva and Volovich.[15] This is obtained by modifying Eqs. (2.12a)–(2.12c) through the introduction of an additional set of superfields $\{Z_t\}_{t=1,\dots,N}$, such that $R(\lambda)$ satisfies (2.12a)–(2.12c), plus the following modified form of (2.12b):

$$(\lambda^2D_{\dot1t}+\lambda Z_t+D_{\dot2t}+A_{\dot2t})R=0.\tag{2.12b′}$$

The integrability conditions for this system are given again by (2.10a)–(2.10c) and (2.11a), together with the following modified form of (2.11b)–(2.11d):

$$D_{\dot1s}(BD_{\dot2t}B^{-1})+D_{\dot1t}(BD_{\dot2s}B^{-1})+\{Z_t,Z_s\}=0,\tag{2.11b′}$$

$$D^s_1(BD_{\dot1t}B^{-1})+2i\delta^s_tA_{1\dot2}+\nabla^s_2Z_t=0,\tag{2.11c′}$$

$$D_{\dot1t}(BD^s_2B^{-1})+2i\delta^s_tA_{2\dot1}+D^s_1Z_t=0,\tag{2.11d′}$$

together with the additional equation

$$\nabla_{\dot\alpha t}Z_2+\nabla_{\dot\alpha s}Z_t=0.\tag{2.11e′}$$

This system reduces to the usual one by setting $Z_t=0$. Unfortunately, it is not relativistically invariant and does not have any clear connection with the supersymmetric field equations. The effect, however, of this modification is to replace Eq. (3.3b) by

$$D_{\dot1t}R_{j+2}+Z_tR_{j+1}+D_{\dot2t}R_j+A_{\dot2t}R_j=0,\tag{3.3b′}$$

which implies, in particular, that instead of $D_{\dot1t}R_1$ vanishing, (3.4f) becomes a definition of Z_t:

$$Z_t=-D_{\dot1t}R_1.\tag{3.4f′}$$

All the following analysis remains unchanged, except that Eq. (3.11d) is no longer necessary as an additional condition and the solutions of the system (4.9a)–(4.9c) given by Theorem 4.4 do in general determine all formal power series solutions of the modified system (2.12a), (2.12b′), and (2.12c).

This, however, only relegates the problem to the determination of appropriate initial data such that the condition $Z_t = 0$ holds; i.e., for which the original constraint equations and supersymmetric field equations are recovered.

V. DISCUSSION

We have shown how to characterize formal power series solutions to the linear system (2.12a)–(2.12c) associated to the constraint equations in terms of a linear infinite-dimensional matrix problem with formal solution given in terms of the Cauchy data by Theorem 4.4. Notice, however, that this is not the Cauchy data for the original problem (2.11), which would involve specification of the superconnection components on the initial data surface plus the imposition of the additional condition (3.11d). In fact, for the $N=3$ supersymmetric field equations, the data should involve fields and their derivatives on a hypersurface in ordinary, rather than super-Minkowski space. The relation between these data and the superconnection components is a rather delicate one involving the elimination of nondynamical parts of the constraint equations and a partial gauge specification eliminating θ-dependent gauge transformations, as described in Ref. 8. Thus, our results cannot be regarded as a solution of the formal power series Cauchy problem in the usual sense, since the initial data $M|_{\mathscr{H}}$ are partly redundant. Nevertheless, all formal power series solutions of the extended system (2.11a) and (2.11b′)–(2.11e′) can be represented in the form given by Theorem 4.4, and this is the precise analog of the solutions determined by Takasaki for the self-dual case in Ref. 18.

As in Ref. 18, it is possible to develop the transformation theory and relate the results to solutions of the Riemann–Hilbert problem. Since the analysis is basically identical, we shall just state the results. Transformations of the type (4.5) map solutions to solutions provided the matrix G is annihilated by the same operators as those in Eqs. (4.13a)–(4.13c) and commutes with Λ. This allows a formal characterization of G in terms of its initial data $G^0 = G|_{\mathscr{H}}$ identical to that of $\widetilde{M}$ in Eq. (4.13). The commutativity with Λ implies an identification, through the map (4.8), with a formal λ-power series expansion

$$g(\lambda) \equiv \sum_{i \in \mathbb{Z}} g_i \lambda^i \leftrightarrow G \equiv \sum_{i \in \mathbb{Z}} g_i \Lambda^i, \tag{5.1}$$

with $g(\lambda)$ satisfying

$$(\lambda D_1^s + D_2^s) g = 0, \tag{5.2a}$$

$$(\lambda^2 D_{\dot{1}t} + D_{\dot{2}t}) g = 0, \tag{5.2b}$$

$$(\lambda^3 \partial_{1\dot{1}} + \lambda^2 \partial_{2\dot{1}} + \lambda \partial_{1\dot{2}} + \partial_{2\dot{2}}) g = 0. \tag{5.2c}$$

It follows that g is actually a function of the following $4 + 2N$ variables (in terms of the coordinates of Sec. IV):

$$g = g(\lambda, v_1 - \lambda(x^{2\dot{2}} + i\theta_t^2 \theta^{\dot{2}t}), v_2 - \lambda^2(v_1 - 2i\theta_t^1 \theta^{\dot{2}t}),$$
$$v_3 - \lambda^2(x^{2\dot{2}} - i\theta_t^2 \theta^{\dot{2}t}),$$
$$\theta_s^1 - \lambda\theta_s^2, \theta^{\dot{1}t} - \lambda^2 \theta^{\dot{2}t}).$$

The resulting $g(x,\theta,\lambda)$ may be regarded as the formal analog of the transition function determining a bundle over a $(4|2N)$-dimensional submanifold of the $(5|2N)$-dimensional superambitwistor space. The Riemann–Hilbert factorization condition

$$g(\lambda) = R_+(\lambda) R_-(\lambda)^{-1},$$

for $R_\pm(\lambda)$ holomorphic in suitably defined coordinate patches, guarantees that these may be regarded as covariant constant sections along super null lines, i.e., solutions of Eqs. (2.12a)–(2.12c) for some globally defined superconnection.

The passage, however, from the formal problem to the holomorphic one requires much deeper analysis. In any case, the difficulties associated with the additional condition (3.11d), and the formal Cauchy problem suggest that to retain full equivalence with the supersymmetric Yang–Mills equations it would be better to return to the original supertwistor formulation of Witten[5] and Manin,[6] based on two spectral parameters.

Note added in proof: In a recent preprint Tafel[22] has reached a similar conclusion regarding the necessity of returning to the full, two spectral parameter linear system, by examining the Riemann–Hilbert factorization problem directly. There are additional cocycle conditions for the patching functions in the general case, however, which do not seem implementable within the present Zakharov–Shabat framework. On the other hand, this fact appears not to present an obstacle within Manin's algebrogeometric approach.[6]

ACKNOWLEDGMENTS

This work was begun while one of us (J.H.) was a visitor at the Institut de Physique Théorique, Université Catholique de Louvain, and continued while the other (M.J.) was visitor at the CRMA, Université de Montréal, both under the "accord d'échange Belgique-Québec," as well as at The Institute for Advanced Study, Princeton where J.H. was a visiting member for the academic year 1984–85. Both authors wish to thank the respective host institutions for hospitality accorded and J. P. Antoine for a careful reading of the preliminary version of the manuscript.

This research was partially supported by National Science Foundation Grant No. MCS-8108814(A04).

[1] S. Mandlestam, "Light-cone superspace and the ultraviolet finiteness of the $N=4$ model," Nucl. Phys. B **213**, 149 (1983).

[2] L. Brink, C. Lindgren, and B. E. W. Nilson, "The ultra-violet finiteness of the $N=4$ Yang–Mills theory," Phys. Lett. B **123**, 323 (1983).

[3] J. H. Schwarz, "Superstring theory," Phys. Rep. **89**, 223 (1982).

[4] M. B. Green, "Supersymmetrical dual string theories and their field theory limits—A review," Surv. High Energy Phys. **3**, 127 (1983).

[5] E. Witten, "An interpretation of classical Yang–Mills theory," Phys. Lett. B **77**, 394 (1978).

[6] Yu. I. Manin, "Flag spaces and supersymmetric Yang–Mills equations," in *Arithmetic and Geometry*, Vol. I, pp. 175–198 of *Progress in Mathematics*, Vol. 35 (Birkhauser, Boston 1983); "New specific solutions and cohomological analysis of ordinary and supersymmetric Yang–Mills equations," Trudy Steklov Inst. **165**, 98 (1983) (in Russian).

[7]I. V. Volovich, "Supersymmetric Yang–Mills equations as an inverse scattering problem," Lett. Math. Phys. **7**, 517 (1983); "Supersymmetric Yang–Mills theories and twistors," Phys. Lett. B **129**, 429 (1983).
[8]J. Harnad, J. Hurtubise, M. Légaré, and S. Shnider, "Constraint equations and field equations in supersymmetric $N=3$ Yang–Mills theory," Nucl. Phys. B **256**, 609 (1985).
[9]M. F. Atiyah, N. J. Hitchin, V. G. Drinfeld, and Yu. I. Manin, "Construction of instantons," Phys. Lett. A **65**, 185 (1978).
[10]N. J. Hitchin, "Monopoles and geodesics," Commun. Math. Phys. **83**, 579 (1982); "On the construction of monopoles," *ibid.* **89**, 145 (1983).
[11]V. E. Zakharov and A. B. Shabat, "A scheme for integrating the nonlinear equations of mathematical physics by the method of the Inverse scattering problem. I and II," Func. Anal. Appl. **8**, 226 (1974) and **13**, 166 (1979).
[12]V. E. Zakharov and A. V. Mikhailov, "Relativistically invariant two-dimensional models of field theory which are integrable by means of the inverse scattering problem method," Sov. Phys. JETP **47**, 1017 (1978).
[13]L. L. Chau, Ge Mo-Lin, and Z. Popowicz, "Riemann–Hilbert transforms and Bianchi–Backlund transformations for the supersymmetric Yang–Mills field," Phys. Rev. Lett. **52**, 1940 (1984).
[14]C. Devchand, "An infinite number of continuity equations and hidden symmetries in supersymmetric gauge theories," Nucl. Phys. B **238**, 333 (1984).
[15]I. Ya. Arefe'eva and I. V. Volovich, "Hidden symmetry algebra for a supersymmetric gauge-invariant model," Lett. Math. Phys. **9**, 231 (1985).
[16]J. Harnad and M. Jacques, "The supersymmetric soliton correlation matrix," J. Geometry Phys. **2**, 1 (1985); "Supersymmetric solitons," in *Group Theoretical Methods in Physics, Proceedings of the XIII International Colloquium,* edited by W. W. Zachary (World Scientific, Singapore, 1984).
[17]S. J. Gates, Jr., M. T. Grisaru, M. Roček, and W. Siegel, *Superspace* (Benjamin/Cummings, Reading, MA, 1983), Chap. 3.
[18]K. Takasaki, "A new approach to the self-dual Yang–Mills equations," Commun. Math. Phys. **94**, 35 (1984).
[19]M. Sato, "Soliton equations as dynamical systems on infinite-dimensional Grassmann manifold," RIMS Kokyuroku **439**, 30 (1981).
[20]J. Harnad, Y. St.-Aubin, and S. Shnider, "The soliton correlation matrix and the reduction problem for integrable systems," Commun. Math. Phys. **93**, 33 (1984).
[21]A. Pressley and G. Segal, *Loop Groups and their Representations* (Oxford U.P., Oxford, 1986).
[22]J. Tafel, "The Riemann–Hilbert problem for the supersymmetric constraint equations," preprint, Bonn-HE-85-33, 1985.

On the formulation of the positive-energy theorem in Kaluza–Klein theories

Osvaldo M. Moreschi
Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 8046 Garching bei München, Federal Republic of Germany
George A. J. Sparling
Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 15 October 1985; accepted for publication 2 April 1986)

The positive-energy theorem is formulated in the context of Kaluza–Klein theories. Different cases are considered, including the situation in which no symmetry is assumed. This work offers a new technique for stability considerations in Kaluza–Klein theories.

I. INTRODUCTION

A few years after the final formulation of the theory of general relativity based on a four-dimensional Riemannian manifold, Kaluza[1] suggested a unified view of gravitation and electromagnetism by means of considering a Riemannian space of five dimensions. In this case just the consideration of a Killing symmetry in the extra dimension introduces a Maxwellian field in the expression of the Ricci tensor.

This initiative was of considerable interest in the following years, resulting in work by Klein[2] and others. Exposition of the ideas of this period can be found in Ref. 3 by Bergmann and Ref. 4 by Lichnerowicz.

Within the framework of Kaluza–Klein theories in five dimensions, Thirring[5] has presented a discussion involving spin-$\frac{1}{2}$ particles, which apparently are provided with an electric dipole moment.

The natural trend of these kinds of ideas led to the study of higher-dimensional spaces in conjunction with non-Abelian gauge fields. This has been the subject of work by DeWitt,[6] Kerner,[7] and Trautman.[8] Later Cho[9] gave a presentation in which the total space P was required to have the structure of a principal fiber bundle; a connection was also given to P and subsequently a metric (i.e., a direct sum of a trivial metric in the fiber and a general metric in the base space M) was introduced in the total space. Under this assumption the symmetry group G turned out to be a group of Killing symmetries for the metric of P.

Salam and Strathdee[10] and Percacci and Ranjbar-Daemi[11] also discussed the case of Killing symmetries, they considered those for which the total space P is given the structure of a bundle over the space-time M with its typical fiber a coset space G/H, instead of a principal bundle.

In Ref. 12 the requirement of a bundle structure for the total space P and of Killing symmetries for the metric g was relaxed, so that only from the condition of conformal symmetries (with possible linear dependence among them) was it shown that there naturally appear Yang–Mills fields, along with Brans–Dicke-like scalars, which contribute to expressions of the Riemann tensor and Ricci tensor and scalar, and so a very general method was obtained for dealing with gravity coupled with these other fields.

Today the belief is widespread that one should not require symmetries on the extra dimensions from the beginning, since they would appear to be just put in by hand. Instead one takes the Ricci scalar of the $(4+n)$-dimensional space as part of the total Lagrangian. Then one supposes that due to some "dynamical mechanism" the "space of minimum energy" will look like $M^0 \times A$, where M^0 is the four-dimensional Minkowski space and A is an n-dimensional compact manifold, which must be characterized by a small parameter in order to be undetected, at normal energies, from the four-dimensional point of view.

Some models[13] have been constructed that illustrate the so-called phenomenon of spontaneous compactification. Unfortunately all such models require the introduction of nongeometrical fields coupled to the curvature tensor, and so one loses the beauty of Kaluza's model, since now there are fields that have to be put in by hand.

Also by mentioning a space of minimum energy one implicitly assumes that one knows what energy means in spaces of higher dimensions; but, as can be seen in the following sections, this is not a trivial question.

There is an endless number of questions to be answered in the consideration of a Kaluza–Klein approach to unification; but one can adopt the attitude that if there is something right about these ideas then it is worthwhile to pursue the study of the classical theory. So this work intends to develop some understanding of the geometrical ideas involved in the Kaluza–Klein approach, by studying the extension of the positive-energy theorems to spaces of higher dimensions.

In the case of the four-dimensional space-time M, positivity is obtained essentially by making use of Stokes' theorem on a hypersurface $\mathbf{N}$, which stretches out to spacelike infinity, to which one attaches an asymptotic boundary $\partial\mathbf{N}$. The idea is that one defines a two-form $\mathbf{E}$, which, when integrated on $\partial\mathbf{N}$, gives the ADM mass[14]; then since Stokes' theorem says that

$$\int_{\partial \mathbf{N}} \mathbf{E} = \int_{\mathbf{N}} \mathbf{dE}, \tag{1}$$

one will get positivity if one can assure that the right-hand side has the appropriate sign.

Using the two-form suggested by Nester,[15] that is,

$$\mathbf{E}_{ab} \equiv \bar{\Psi}\gamma_{[a}\mathbf{D}_{b]}\gamma_5\Psi + \text{b.c.}\,; \tag{2}$$

where b.c. means "bar conjugate," the ADM momentum is determined by

 0022-2488/86/092402-07$02.50

$$\int_{\partial N} \mathbf{E} = -8\pi P^a_{\rm ADM}\, \overline{\Psi}_0 \gamma_a \Psi_0 , \tag{3}$$

as long as the Dirac spinor Ψ behaves as $\Psi = \Psi_0 + O(1/r)$, where Ψ_0 is a covariantly constant spinor at spacelike infinity.

The three-form $\mathbf{dE}$ under the integral sign can be given by

$$\mathbf{dE}_{\underline{abc}} = (1/3!)\epsilon_{0\underline{abc}}\,[2\mathbf{D}^{\underline{e}}(\overline{\Psi})\gamma^0 \mathbf{D}_{\underline{e}}(\Psi) + 2\mathbf{D}_{\underline{e}}(\overline{\Psi})\gamma^{\underline{e}}\gamma^0\gamma^{\underline{d}}\mathbf{D}_{\underline{d}}\Psi + \mathbf{G}^{0\underline{e}}\overline{\Psi}\gamma_{\underline{e}}\Psi] ; \tag{4}$$

where $\underline{a}, \underline{b},... = 1,2,3$ refer to the spacelike components of an orthonormal tetrad in which the timelike vector is orthogonal to N.

The Einstein equation, with the conventions used here $[G = c = 1, \mathbf{g} = (+---)]$, is

$$\mathbf{G}_{ab} = -8\pi\, T^T_{ab} , \tag{5}$$

where T^T_{ab} is the "total" energy-momentum tensor. Then one observes that the dominant energy condition, along with the Einstein equation, implies that the third term in Eq. (4) is negative.

By imposing the equation

$$\gamma^{\underline{d}}\mathbf{D}_{\underline{d}}\Psi = 0 , \tag{6}$$

on the hypersurface N, one obtains that the whole expression in Eq. (4) is negative, which implies that the ADM momentum is a future-directed timelike vector.

This approach to the positivity of the ADM mass relies then on the possibility of having solutions of Eq. (6) with the behavior $\Psi = \Psi_0 + O(1/r)$—a fact that already has been established.[16]

It is also possible to formulate a theorem in which it is proved that the energy determined at null infinity, that is, the Bondi mass,[17] is positive.[18] This result has an interesting physical content, since it assures us that if an isolated system, that has at certain retarded time u mass $m(u)$, is losing energy in the form of gravitational waves, then energy lost in the future of u cannot exceed $m(u)$, since the Bondi mass cannot become negative.

The definitions of the ADM and Bondi mass are stated for spaces that are asymptotically flat at spacelike and null infinity, respectively. To fix ideas, from now on "asymptotically flat" will be meant at spacelike infinity in the sense of Ref.[19]; although it should be remarked that one needs much weaker conditions for the theorem to go through, as is discussed in the second paper of Ref. 16.

The positivity of mass is expected from a theory that intends to describe attractive gravitational interactions, since, for example, test particles are repelled by compact objects with negative mass.

The description of physical phenomena is up to now carried out by making use of the notion of a smooth four-dimensional space-time. In the Kaluza–Klein approach it is conjectured that actually there are other spacelike dimensions that we have not detected yet. The idea is that at every point of the normal four-dimensional space-time M there is an "internal" space that extends M to a $(4+n)$-dimensional space-time P. The principles of general relativity in M then require that the internal space at any point of M should be (at least) topologically the same; in other words, P should be a fiber bundle over M. If the "real" space-time is actually P, then one wants to see that this description is not incompatible with our present experience. In particular one can ask: Are the effective gravitational interactions detected in M still attractive? In answering this question one immediately is led to a formulation of the positive energy theorem in P.

In the following sections the spinorial techniques used for the case of the four-dimensional space-time M will be extended to the case of a higher-dimensional space-time P. That is, a form E is defined in a way that gives the energy momentum vector of P. Then applying Stokes' theorem one studies the exterior derivative of E, which will be given by expressions analogous to Eq. (4). The appearance in these expressions of the corresponding Einstein tensor of the space-time P is particularly significant due to the fact that it is still not known which are the "correct" field equations in the Kaluza–Klein approach, so the treatment developed in the following sections will be done without using the field equations for the Einstein tensor in the $(4+n)$-dimensional space-time P. On the contrary, one believes that one knows the appropriate field equations in the four-dimensional space-time M; so they will be assumed to hold. In this way the expressions analogous to Eq. (4) will have terms involving the matter fields, which will be subject to some conditions in order to obtain the desired positivity.

In Sec. II, the space P will be considered to be a five-dimensional space-time with a cyclic symmetry.

The case in which the total space is a non-Abelian principal fiber bundle over the four-dimensional space-time M is treated in Sec. III.

The more general case in which P is not required to have global symmetries will be discussed in Sec. IV. This treatment is also applicable to cases in which P has symmetries, but which have not been covered in previous sections.

Although the following results will be concerned with the notion of mass at spatial infinity (ADM mass), everything can be extended to consider the notion of mass at null infinity (Bondi mass), by for example, using the techniques described in the last paper of Ref. 18.

Objects defined on the four-dimensional space-time M will be normally in boldface, in order to distinguish them from similar objects defined on P.

II. POSITIVE-ENERGY THEOREMS IN 4 +1 DIMENSIONS

In this section, the total space P is taken to be a five-dimensional manifold with metric g, which has signature -3 and a global cyclic Killing symmetry K_1 (K_1 is ∇_1 of Ref. 12). Furthermore P is assumed to have the structure of a bundle over the four-dimensional manifold M with fiber S^1. The integral lines of the vector field K_1 are the fibers of P.

The metric g can be expressed by

$$g = g_{ab}\theta^a \otimes \theta^b + F\theta^1 \otimes \theta^1 , \tag{7}$$

where the notation developed in Ref. 12 is being used. The indices $a, b,...,$ which here label tensors in the orthogonal space to K_1, are the same abstract indices used to label tensors on M. In Eq. (7) the abstract indices that would remark the tensor character of the metric g are being omitted for reason of simplicity. In this way θ^a is a one-form in P with an extra index "a"; in particular it explicitly gives the pull-back of a one-form $\mathbf{w}_a$ in M to a one-form $\pi^*(\mathbf{w})$ in P by $\mathbf{w}_a\theta^a = \pi^*(\mathbf{w})$, where π is the projection map in the bundle P. The function F is the scalar product of K_1 with itself.

Since g_{ab} is Lie-derived by the Killing vector K_1, it is the lift of a metric $\mathbf{g}$ from the base manifold M. The space-time $(M,\mathbf{g})$ is assumed to be asymptotically flat.

Let N be a spacelike hypersurface in M, which stretches out to spacelike infinity, with asymptotic boundary ∂N. The inverse image under the projection π of N to the space P defines the hypersurface N with boundary ∂N.

The appropriate three-form E to be integrated on ∂N is

$$E_{\mathrm{ABC}} \equiv \tfrac{1}{2}\,\overline{\Psi}\gamma_{[\mathrm{A}}\gamma_{\mathrm{B}}\gamma D_{\mathrm{C}]}\Psi + \text{b.c.}\,, \tag{8}$$

where now A,B,C,...are abstract indices for tensors in P; and the elements γ_{A} are the generators of the Clifford algebra in P, that is,

$$\gamma_{\mathrm{A}}\gamma_{\mathrm{B}} + \gamma_{\mathrm{B}}\gamma_{\mathrm{A}} = 2g_{\mathrm{AB}}\,I\,,$$

where I is the identity element of the algebra. Also γ is given by

$$\gamma \equiv (1/5!)\epsilon^{\mathrm{ABCDE}}\,\gamma_{\mathrm{A}}\gamma_{\mathrm{B}}\gamma_{\mathrm{C}}\gamma_{\mathrm{D}}\gamma_{\mathrm{E}}\,, \tag{9}$$

where ϵ is the volume form, and Ψ belongs to a space of representation of the Clifford algebra, that is, Ψ is a spinor. The bar operation is defined analogously as in the case of four dimensions. (A detailed study of Clifford algebra representations, spinors, spin connection, and Lie derivative of spinors is found in Ref. 20.)

Assuming the spinor Ψ behaves as $\Psi_0 + O(1/r)$, where Ψ_0 is a covariantly constant spinor at ∂N and where $1/r$ can be taken as $\Omega^{1/2}$ of Ref. 19, then the integral of the three-form E on ∂N gives

$$\int_{\partial N} E = -8\pi\Omega\overline{\Psi}_0(p^a_{\mathrm{ADM}}\gamma_a + sQ_E\gamma_5 - sQ_M)\Psi_0\,, \tag{10}$$

where Ω is the length of the fiber, s is a free parameter, and Q_E and Q_M are the electric and magnetic charge of the electromagnetic field F_{ab}. To obtain Eq. (10) it was assumed that $F = -1 + O(1/r)$ for large values of r, and also that

$$R_{ab}{}^1 = 4sF_{ab}\,, \tag{11}$$

where $R_{ab}{}^1$ is $R_{ab}{}^\alpha$ of Ref. 12 for $\alpha = 1$, which is the curvature of the connection in the principal fiber bundle P defined by the orthogonal complement of the Killing symmetry.

From Stokes' theorem one obtains

$$\int_{\partial N} E = \int_N dE\,; \tag{12}$$

so, similarly as was done in the case of four dimensions, one wants to study the expression of dE under the integral sign, which is given by

$$dE_{\underline{ABCD}} = (1/4!)\epsilon_{\underline{ABCD}}\,[2D^{\underline{E}}(\overline{\Psi})\gamma^0 D_{\underline{E}}(\Psi) + 2D_{\underline{E}}(\overline{\Psi})\gamma^{\underline{E}}\gamma^0\gamma^{\underline{F}}D_{\underline{F}}\Psi + G^{0\underline{E}}\overline{\Psi}\gamma_{\underline{E}}\Psi]\,, \tag{13}$$

where $\underline{A},\underline{B}$,... refer to the spacelike components of an orthonormal basis in which the timelike vector is orthogonal to the hypersurface N. Note that Eq. (13) is the exact analog of Eq. (4); in particular now D_{A} is the Riemann connection in P, and G_{AB} is the Einstein tensor of the metric g of P.

Then from Eq. (12), to get a statement about a definite sign for the right-hand side of (10), one has to require a definite sign for (13), which in particular forces one to make a decision about what to do with the third term of Eq. (13). The essential difference with the case of four dimensions is that, while many people agree about using the Einstein equation for the Einstein tensor in four dimensions $\mathbf{G}_{ab}$, there is no general agreement on what should be the field equation for G_{AB}. To get some insight, in the rest of this section two different situations will be considered.

Case (a): Assuming $F = -1$, one wants to mimic the conditions imposed in the four-dimensional case; namely to ask for

$$\gamma^{\underline{A}}D_{\underline{A}}\Psi = 0 \tag{14}$$

and

$$G^{0\mathrm{A}}\overline{\Psi}\gamma_{\mathrm{A}}\Psi \leqslant 0\,, \tag{15}$$

which then will imply that the expression (13) is negative.

But, what are the allowed values of s? To answer this question one has to look more closely at G_{AB}.

In this case one has

$$G_{ab} = \mathbf{G}_{ab} + 8\pi(2s)^2T^{\mathrm{EM}}_{ab}\,, \tag{16}$$

$$G_{a\alpha} = 2s\nabla^b F_{ba}\,,\quad \alpha = 1\,, \tag{17}$$

where

$$T^{\mathrm{EM}}_{ab} \equiv -(1/4\pi)(F_a{}^cF_{bc} - (g_{ab}/4)F_{cd}F^{cd}) \tag{18}$$

is the energy-momentum tensor of the electromagnetic field. It should be emphasized that Eqs. (16) and (17) are not field equations, they just represent the calculation of certain "components" of the Einstein tensor in five dimensions assuming the identification (11) with $F = -1$.

It is reasonable to require that the appropriate field equation for G_{AB} should have as a consequence the standard Einstein–Maxwell equations on the four-dimensional space-time M, namely

$$\mathbf{G}_{ab} = -8\pi(T_{ab} + T^{\mathrm{EM}}_{ab})\,, \tag{19}$$

$$\nabla_{[a}F_{bc]} = 0\,, \tag{20}$$

$$\nabla^b F_{ba} = 4\pi J_a\,, \tag{21}$$

where T_{ab} is the energy-momentum tensor of matter.

If one, on the contrary, substitutes these equations in (16) and (17), one gets

$$G_{ab} = -8\pi\{T_{ab} + [1-(2s)^2]T^{\mathrm{EM}}_{ab}\}\,, \tag{22}$$

$$G_{a\alpha} = 8\pi s\,J_a\,,\quad \alpha = 1\,. \tag{23}$$

Then one observes that, since $T^{\mathrm{EM}}_{ab}u^av^b \geqslant 0$ for any two future-directed timelike vectors u^a and v^b, one will get condition (15) imposing

$$T_{ab}u^av^b \geqslant |s\,J_a u^a| \tag{24}$$

for any two future-directed timelike unit vectors u^a and v^b, along with the condition

$$|s| \leqslant \tfrac{1}{2}\,. \tag{25}$$

Summarizing, one gets that conditions (14), (24), and (25) imply that (10) is negative, which has a consequence that the ADM mass satisfies

$$m \geqslant |s| \sqrt{Q_E^2 + Q_M^2}\,. \tag{26}$$

This result cannot be the optimum result, since one knows of the theorem of Gibbons and Hull,[21] where they get (26) but with $|s|$ replaced by 1.

It is interesting to note that for the maximum allowable value of $|s|$ in this case, i.e., $|s| = \frac{1}{2}$, Eqs. (22) and (23) take the form

$$\text{geometry} = \text{matter}\,; \tag{27}$$

so in particular, when there is no matter, one would have equations involving only "geometry" that couple gravity and electromagnetism.

Case (b): Here one wants to change the conditions so that one gets inequality (26), but with $|s|$ not bounded by $\frac{1}{2}$.

So, again consider the case $F = -1$, and assume the Einstein–Maxwell equations (19)–(21) to hold.

Now one defines

$$\widetilde{D}_{\mathrm{A}}\Psi \equiv D_{\mathrm{A}}\Psi + \gamma_{\mathrm{A}}(s/4)F^{ab}\gamma_a\gamma_b\gamma_5\Psi\,, \tag{28}$$

and the Lie derivative of spinors[20] in the direction of a vector v by

$$\mathscr{L}_v\Psi = D_v\Psi + \tfrac{1}{4}D^{[\mathrm{A}}v^{\mathrm{B}]}\gamma_{\mathrm{A}}\gamma_{\mathrm{B}}\Psi\,, \tag{29}$$

$$\mathscr{L}_v\gamma_{\mathrm{A}} = D_{(\mathrm{A}}v_{\mathrm{B})}\gamma^{\mathrm{B}}\,. \tag{30}$$

Then one observes that by imposing the conditions

$$\mathscr{L}_{K_1}\Psi = 0\,, \tag{31}$$

$$\gamma^{g}\widetilde{D}_{g}\Psi = 0\,, \tag{32}$$

one obtains

$$\begin{aligned} dE_{\underline{ABCD}} = (1/4!)\epsilon_{\underline{ABCD}0}\{&2\widetilde{D}^{g}(\overline{\Psi})\gamma^0\widetilde{D}_{g}(\Psi) \\ &- 8\pi(T^{0e} + (1-s^2)T^{0e}_{\mathrm{EM}})\overline{\Psi}\gamma_e\Psi \\ &+ sJ^0\overline{\Psi}\gamma_5\Psi]\}\,. \end{aligned} \tag{33}$$

One requires condition (32) by applying a technique described in Ref. 22 that maximizes the allowable values of s.

It then can be deduced easily that the additional condition (24), now with $|s| \leqslant 1$, assures that the expression (33) is negative, which implies the following result.

Theorem: Let P be a five-dimensional manifold with metric g that has signature -3 and a global cyclic Killing symmetry K_1, such that $g(K_1,K_1) = -1$.

Let P also have the structure of a bundle over the four-dimensional manifold M with fiber S^1, which are the integral lines of the vector field K_1.

Under these conditions then the restriction of the metric g to the orthogonal complement of K_1 is the lift to P of some metric $\mathbf{g}$ from M. The space-time $(M,\mathbf{g})$ is assumed asymptotically flat, and the Einstein–Maxwell field equations (19)–(21) are satisfied in M.

Then the condition on the matter field

$$T_{ab}u^a v^b \geqslant |s\,J_a u^a|\,, \tag{24}$$

for any two future-directed timelike unit vectors u^a and v^b, and with

$$|s| \leqslant 1\,, \tag{34}$$

implies that the corresponding energy-momentum vector of P, that is, $(P^a_{\mathrm{ADM}}, sQ_E)$, is nonspacelike and future directed and also implies that

$$m \geqslant |s| \sqrt{Q_E{}^2 + Q_M{}^2}\,. \tag{35}$$

This result can be completely restated in terms of objects intrinsic to the four-dimensional space-time M. In particular it has been shown in Ref. 22 that since there is an experimental bound to the ratio of charge density to mass density given by $|e/m_e|$, where "e" and "m_e" are the charge and mass of the electron, then one can find for each physical system a value s_0 such that the matter condition is automatically satisfied.

III. POSITIVITY OF ENERGY IN A PRINCIPAL FIBER BUNDLE OVER THE SPACE-TIME *M*

Let P now be $(4+n)$-dimensional manifold with metric g, which has signature $-(n+2)$ and admits n pointwise linearly independent Killing vector fields K_α, $\alpha = 1,\ldots,n$ (the K_α are the $\mathbf{V}_\alpha$ of Ref. 12). Let these vector fields form a Lie algebra of a semisimple compact Lie group G. Let $K_{\alpha\beta}$ be the negative definite Killing–Cartan form of G. Then it will be assumed that

$$g(K_\alpha, K_\beta) = K_{\alpha\beta}\,. \tag{36}$$

Furthermore let P have the structure of a principal fiber bundle over the four-dimensional manifold M, where the vector fields K_α span the tangent space of the fibers at each point of P.

Similarly as was done in the previous section, the metric g can be expressed[12] by

$$g = g_{ab}\theta^a \otimes \theta^b + K_{\alpha\beta}\theta^\alpha \otimes \theta^\beta\,. \tag{37}$$

Since g_{ab} is Lie-derived by the Killing vectors K_α, it is the lift of a metric $\mathbf{g}$ from the base manifold M. The space-time $(M,\mathbf{g})$ is assumed to be asymptotically flat. A space with all these characteristics will be denoted by $P(g,G,K_{\alpha\beta},M)$.

As was done before, the spacelike hypersurface N, of P, with boundary ∂N, is constructed by the inverse image, under the projection of the bundle, of the spacelike hypersurface N, of M, which stretches out to spacelike infinity and has asymptotic boundary $\partial\mathrm{N}$.

The $(2+n)$-form E, whose integral on ∂N will give the energy-momentum vector of the space-time P, is the natural extension of the one defined by Eq. (8), namely,

$$E_{\mathrm{A}_1\cdots\mathrm{A}_{2+n}} \equiv [1/(n+1)!]\overline{\Psi}\gamma_{[\mathrm{A}_1}\cdots\gamma_{\mathrm{A}_{1+n}}\gamma\mathscr{D}_{\mathrm{A}_{n+2}]}\Psi + \text{b.c.}\,, \tag{38}$$

where γ_{A} is a generator of the Clifford algebra in the space-time (P,g),

$$\gamma \equiv [1/(4+n)!]\epsilon^{\mathrm{A}_1\cdots\mathrm{A}_{4+n}}\gamma_{\mathrm{A}_1}\cdots\gamma_{\mathrm{A}_{4+n}}\,, \tag{39}$$

and Ψ as before is a spinor in P.

The connection $\mathscr{D}$ cannot be taken to be the Riemann connection when G is a non-Abelian Lie group. This is so because following the model used in previous sections, one wants to require the spinor Ψ to be asymptotically constant; which implies

$$\mathscr{D}_{\mathrm{A}}\Psi|_{\infty}=0\,. \tag{40}$$

But this condition at spacelike infinity requires that the curvature of the connection $\mathscr{D}$ has to be zero there, which is not the case of the Riemann connection of the metric g.

There are two natural metric connections in this case[23] that will do the job; in this section $\mathscr{D}$ will be taken to be the metric connection with torsion

$$T=c_{\alpha\beta}{}^{\sigma}\theta^{\alpha}\otimes\theta^{\beta}\otimes\theta K_{\sigma}\,, \tag{41}$$

where $c_{\alpha\beta}{}^{\sigma}$ are the structure constants of the Lie algebra generated by the K_α's, that is,

$$[K_\alpha,K_\beta]=c_{\alpha\beta}{}^{\sigma}K_\sigma\,. \tag{42}$$

Then if the spinor Ψ behaves as $\Psi_0+O(1/r)$, where Ψ_0 is a covariantly constant spinor at ∂N, the integral of the $(2+n)$-form E on ∂N gives

$$\int_{\partial N}E=-8\pi\Omega(-1)^{n(n+1)/2}\times\overline{\Psi}_0(P_{\mathrm{ADM}}{}^{a}\gamma_a+sQ_E{}^{a}\gamma_a+sQ_M{}^{a}\gamma_5\gamma_a)\Psi_0\,, \tag{43}$$

where the electric and magnetic Yang–Mills charges can be defined by

$$8\pi\,Q_E^{\alpha}\,\overline{\Psi}_0\gamma_a\Psi_0=-\int_{\partial\mathrm{N}}\overline{\Psi}{*}F^{\alpha}\gamma_a\Psi \tag{44}$$

and

$$8\pi\,Q_M^{\alpha}\,\overline{\Psi}_0\gamma_5\gamma_a\Psi_0=\int_{\partial\mathrm{N}}\overline{\Psi}F^{\alpha}\gamma_5\gamma_a\Psi\,, \tag{45}$$

Ω denotes the volume of the fiber, and s is a free parameter that relates the geometric object $R_{ab}{}^{\alpha}$ defined in Ref. 12 with the Yang–Mills field $F_{ab}{}^{\alpha}$ at the space-time M by the equation

$$R_{ab}{}^{\alpha}=4s\,F_{ab}{}^{\alpha}\,. \tag{46}$$

In analogy with the four-dimensional case, the energy-momentum vector of the total space-time P is defined by $(P^{a}_{\mathrm{ADM}},s\,Q_E{}^{\alpha})$; the aim is then to establish that it is nonspacelike and future directed.

In applying again Stokes' theorem [Eq. (12)], one wants to study the spacelike components of the exterior derivate dE of the form E defined by (38). Here it is convenient to make use of the experience of the last section, so below one will follow the model of case (b).

The Einstein–Yang–Mills equations are assumed to hold, that is,

$$\mathbf{G}_{ab}=-8\pi(T_{ab}+T^{\mathrm{YM}}_{ab})\,, \tag{47}$$

where

$$T^{\mathrm{YM}}_{ab}=(1/4\pi)(F_a{}^{c\sigma}F_{bc\sigma}-\tfrac{1}{4}g_{ab}F_{cd\sigma}F^{cd\sigma})\,, \tag{48}$$

and also

$$\nabla_bF^{ba\beta}=4\pi J^{a\beta}\,, \tag{49a}$$

$$\nabla_{[a}F_{bc]}{}^{\beta}=0\,. \tag{49b}$$

Equation (49b) is actually a consequence of (46). Defining

$$\widetilde{D}_{\mathrm{A}}\Psi\equiv\mathscr{D}_{\mathrm{A}}\Psi+\gamma_{\mathrm{A}}(s/4)F^{ab\alpha}\gamma_a\gamma_b\gamma_\alpha\,\Psi\,, \tag{50}$$

one requires

$$\gamma^{\underline{a}}\widetilde{D}_{\underline{a}}\Psi=0\,. \tag{51}$$

Also it is assumed that the spinor are Lie-derived in the direction of the Killing symmetries, that is,

$$\mathscr{L}_{K_\alpha}\Psi=0\,. \tag{52}$$

It can be proved that the integrability conditions of Eq. (52) are automatically satisfied, and that (52) is compatible with (51).

Then using Eqs. (47), (49), (51), and (52), and after a very long calculation,[20] one obtains

$$dE_{\underline{A}_1\cdots\underline{A}_{n+3}}=\frac{(-1)^{n(n+1)/2}}{(n+3)!}\epsilon_{0\underline{A}_1\cdots A_{n+3}}[2\,\widetilde{D}^{\underline{a}}\overline{\Psi}\gamma^0\widetilde{D}_{\underline{a}}\Psi-8\pi(T^{0b}+(1-s^2)\,T^{0b}_{\mathrm{YM}})\overline{\Psi}\gamma_b\Psi-8\pi\,s\,J^{0\beta}\overline{\Psi}\gamma_\beta\Psi+s\,\nabla_\sigma(F_{ab\delta})\overline{\Psi}\gamma^{[0}\gamma^a\gamma^{b]}\gamma^\sigma\gamma^\delta\Psi]\,. \tag{53}$$

Since the first term is negative, and the term involving the energy-momentum tensor of the Yang–Mills field is negative, too (for $|s|\leqslant 1$), it is only left to impose some condition on the matter tensor T_{ab} in order to get the desired result. This condition is

$$T^{ab}u_a\overline{\Psi}\gamma_b\Psi\geqslant|s|\overline{\Psi}(J^{a\beta}u_a\gamma_\beta-(1/8\pi)\nabla_\sigma(F_{ab\delta})\gamma^{[c}\gamma^a\gamma^{b]}\gamma^\sigma\gamma^\delta u_c)\Psi\,, \tag{54}$$

for any future-directed timelike vector u^a. One then has the following result.

Theorem: Let the $(4+n)$-dimensional space-time $P(g,G,K_{\alpha\beta},M)$ be defined as at the beginning of this section.

Then condition (54) on the matter fields implies that the energy-momentum vector of P, defined at spatial infinity, is nonspacelike and future directed, and also implies that the ADM mass satisfies

$$m\geqslant|s|\sqrt{(Q_E^{\alpha})^2+(Q_M^{\alpha})^2}\quad\text{(summed over }\alpha) \tag{55}$$

with

$$|s|\leqslant1\,.$$

In (55) one has assumed without loss of generality that the Killing–Cartan form $K_{\alpha\beta}$ is minus the Kronecker delta.

It is then observed that to get the nice result expressed by inequality (55), one has to introduce the second term on the right-hand side of (54). The physical significance of this term is not clear yet: it is an effect of the non-Abelian character of the Lie group G. On the contrary, the first term on the right-hand side of (54) is expected, and in fact it is the analog of the one appearing in Eq. (24).

It is important to note that Eq. (52) determines the behavior of the spinor along the fibers, so that Eq. (51) is effectively an equation on the hypersurface $\mathbf{N}$ of M, which then permits[24] us to use the arguments of Ref. 16 to prove the existence of solutions of (51).

To consider the case in which black holes are present in the interior of the space-time M, one has to use the procedure described in Ref. 25.

IV. POSITIVITY WITHOUT SYMMETRIES?

Here one wants to formulate the positive-energy theorem for a space P that is not assumed to have global symmetries.

As was mentioned already the minimum that is required in a Kaluza–Klein description of space-time is that at any point of our experienced four-dimensional space-time M, there is an internal space of n dimension that extends M to a total space P. In other words the minimum physical assumption on P is that it has the structure of a bundle over M with fiber B. It is generally believed that B should be a compact space.

As in the case of four dimensions one needs the notion of asymptotic flatness in order to be able to formulate the positive-energy theorem. The least that has to be understood by asymptotic flatness is that the metric g of P can be expressed by

$$g = g^0 + O(1/r) , \tag{56}$$

where g^0 has Killing symmetries $K_{\alpha'}$, $\alpha' = 1,...,p$, $p \geqslant n$, which spans the fiber B asymptotically. Furthermore g^0 is supposed to be expressible by

$$g^0 = g_{ab}\theta^a \otimes \theta^b + g_{\alpha\beta}\theta^\alpha \otimes \theta^\beta , \tag{57}$$

where g_{ab} is the lift of a flat metric $\mathbf{g}^0_{ab}$ from M. As before the function r can be identified with $\mathbf{r}$ from M.

It is clear then that this notion of asymptotic flatness does not mean that P is flat at infinity since there is no condition on the curvature of the internal space. Instead one should look at this notion as the least one expects from a space P that describes an isolated system, i.e., an isolated system from the M point of view.

Next one needs to have a $(n+2)$-form E such that when it is integrated at infinity it gives the corresponding energy-momentum vector. As before, one takes E to be

$$E_{\mathsf{A}_1 \cdots \mathsf{A}_{n+2}} = [1/(n+1)!]\overline{\Psi}\gamma_{[\mathsf{A}_1} \cdots \gamma_{\mathsf{A}_{n+1}}\gamma\mathscr{D}_{\mathsf{A}_{n+2}]}\Psi + \text{b.c.} , \tag{38}$$

where $\mathscr{D}$ also has to have zero curvature at infinity. Then for spinors Ψ that are asymptotically constant, the energy-momentum P^{A} of P is defined by

$$\int_{\partial N} E = -8\pi P^{\mathsf{A}}\overline{\Psi}_0\gamma_{\mathsf{A}}\Psi_0 . \tag{58}$$

Here one touches the most critical point in this procedure, since even if one has a definite prescription on how to calculate $\mathscr{D}$ at infinity, there will be in general no canonical way of extending $\mathscr{D}$ to the interior of P, which is necessary for the application of Stokes' theorem. In other words there will be a high degree of arbitrariness in the calculation of the exterior derivative dE; and so if one wants to require its integral on the hypersurface N to have a definite sign, this necessarily will involve an arbitrary condition on the Einstein tensor G_{AB} of P. This condition depends on how one extends the connection D to the interior of P.

This difficulty makes it very hard to give any physical meaning to positivity of energy when the space (P,g) has only the structure of a bundle in the interior. If, on the contrary, g has global Killing symmetries, $\mathscr{D}$ naturally can be extended to the interior and this difficulty disappears.

In what follows then, one has in mind two cases: one in which no symmetries are assumed at all, and the other will be the spaces with symmetries not treated in the previous sections.

By the requirement of $\mathscr{D}$ to be a metric connection, one knows that it is completely determined by its torsion $T_{\mathsf{AB}}{}^{\mathsf{C}}$. One can express the exterior derivative dE in terms of $\mathscr{D}$ and T

$$dE_{\mathsf{A}_1 \cdots \mathsf{A}_{n+3}} = \mathscr{D}_{[\mathsf{A}_1}E_{\mathsf{A}_2 \cdots \mathsf{A}_{n+3}]} + [(n+2)/2]T_{[\mathsf{A}_1\mathsf{A}_2}{}^{\mathsf{B}}E_{|\mathsf{B}|\mathsf{A}_3 \cdots \mathsf{A}_{n+3}]} . \tag{59}$$

This expression suggests taking the torsion such that $T_{\mathsf{AB}}{}^0 = 0$, where 0 is the direction orthogonal to the spacelike hypersurface N. By doing this one obtains that

$$dE_{\underline{\mathsf{A}}_1 \cdots \underline{\mathsf{A}}_{n+3}} = \frac{(-1)^{n(n+1)/2}}{(n+3)!}\epsilon_{0\underline{\mathsf{A}}_1 \cdots \underline{\mathsf{A}}_{n+3}} \times \{2\mathscr{D}_{\underline{\mathsf{A}}}\overline{\Psi}\gamma^{\underline{\mathsf{A}}}\gamma^0\gamma^{\underline{\mathsf{B}}}\mathscr{D}_{\underline{\mathsf{B}}}\Psi + 2\mathscr{D}^{\underline{\mathsf{A}}}\overline{\Psi}\gamma^0\mathscr{D}_{\mathsf{A}}\Psi + \Psi\gamma^{[0}\gamma^{\mathsf{B}}\gamma^{\mathsf{D}]}[\tfrac{1}{8}\mathbb{R}_{\mathsf{CD}}{}^{\mathsf{EF}}\gamma_{\mathsf{E}}\gamma_{\mathsf{F}}\Psi + T_{\mathsf{CD}}{}^{\mathsf{E}}\mathscr{D}_{\mathsf{E}}\Psi]\} . \tag{60}$$

One can always express the curvature $\mathbb{R}$ by

$$\mathbb{R}_{\mathsf{ABC}}{}^{\mathsf{D}} = R_{\mathsf{ABC}}{}^{\mathsf{D}} + \hat{R}_{\mathsf{ABC}}{}^{\mathsf{D}} , \tag{61}$$

where R is the Riemann curvature and $\hat{R}$ depends on the torsion T and its derivatives. Then the last two terms in Eq. (60) look like

$$G^{0\mathsf{B}}\overline{\Psi}\gamma_{\mathsf{B}}\Psi + \overline{\Psi}\Delta^0\Psi , \tag{62}$$

where Δ^{A} is an operator that depends on T and its derivatives. Then one observes that by requiring

$$\gamma^{\underline{\mathsf{A}}}\mathscr{D}_{\underline{\mathsf{A}}}\Psi = 0 \tag{63a}$$

and

$$-G^{\mathsf{AB}}u_{\mathsf{A}}\overline{\Psi}\gamma_{\mathsf{B}}\Psi \geqslant |u_{\mathsf{A}}\overline{\Psi}\Delta^{\mathsf{A}}\Psi| , \tag{63b}$$

for any future-directed timelike vector u, one concludes that the energy-momentum vector defined by E in Eq. (58) is timelike and future pointing.

One should note that the last term in Eq. (60) involves derivatives of the spinor and so does Δ^{A} in inequality (63b). It is convenient then to use the same technique as in the case discussed in the previous section, namely to fix the Lie derivative of the spinor in the vertical direction. This again naturally can be done if one has available global Killing symmetries on P.

V. FINAL COMMENTS

It may appear that the cases treated in Secs. II and III are too special, but actually they may be the most important situations to be considered, since they make use of the minimum number of fields necessary to describe a nontrivial Kaluza–Klein space-time.

In case (a) of Sec. II it was observed that a condition as natural as (15) leads to an inequality where the parameter s, described before, is constrained by the condition $|s| \leqslant \frac{1}{2}$. In case (b) of Sec. II it was shown that one can obtain a better inequality by relaxing the condition on the Einstein tensor of P, and by imposing conditions directly on the matter fields instead. It can also be seen, by methods similar to those of Ref. 26, that zero mass implies that P is flat. It is worthwhile to remark that "flat" in this case does not mean five-dimensional Minkowski space, since one of the spacelike dimensions is compact. Therefore considerations of stability in Kaluza–Klein space-times are meaningless unless one refers to some specific topology. In fact, Witten has speculated[27] that the $M_0 \times S^1$ five-dimensional Kaluza-Klein space-time might be unstable against "a process of semiclassical barrier penetration." Unfortunately, it is still not known what is the appropriate theory that incorporates quantum and gravitational effects on the same footing, and so one really does not know yet how to treat the stability properties of a space-time that come from quantum effects. In any case, one can consider the example given in Ref. 27 by Eq. (6) and ask what is the assumption of the positive-energy theorem in five dimensions that is not satisfied by it? The answer is that this example does not have the structure of a bundle over a four-dimensional space-time M with fiber S^1. Even if one concentrates on the asymptotic region, the example of Ref. 27 fails to be asymptotically flat at null infinity since one cannot construct a complete scri.

Section III shows the convenience and necessity of introducing a torsion metric connection when dealing with non-Abelian gauge fields. The case treated in this section also points out the kind of complicated conditions that have to be required on the matter fields when one starts to consider more general situations.

In Sec. IV, the general procedure for the positive-energy theorem in the Kaluza–Klein context has been formulated, without assuming a particular field equation or the existence of symmetries. If one has more structure, and if one knows the field equation, then conditions (63) are supposed to be changed in order to adapt them to the available structure. Assuming the wish of having a positive-energy theorem for a Kaluza–Klein theory, the resulting conditions on the matter field can be used to measure how physically reasonable the field equations are; since conditions that are physically difficult to understand should be avoided.

Finally, the techniques described above offer a new possibility in considerations of classical stability in Kaluza–Klein theories, since for a complete theory (that is, where the field equations are known), one can apply the methods used for the case of the standard four-dimensional space-time M discussed in Ref. 26, in order, for example, to study the manifold of solutions with zero energy.

ACKNOWLEDGMENTS

One of us (O. M. M.) would like to thank the hospitality of the General Relativity Group at the Max-Planck-Institut für Astrophysik, and D. Brill and O. Reula for stimulating and clarifying discussions in the subject of Kaluza–Klein theories. Also we would like to express our gratitude to B. Schmidt, who suggested several improvements in the manuscript.

[1] T. Kaluza, Sitzungsber. Preuss. Akad. Wiss. **1921**, 966.
[2] O. Klein, Z. Phys. **37**, 859 (1926).
[3] P. G. Bergmann, Int. J. Theor. Phys. **1**, 25 (1968).
[4] A. Lichnerowicz, *Theories relativistes de la gravitation et de l'electromagnetisme* (Masson, Paris, 1955).
[5] W. Thirring, in *11th Schladming Conference*, edited by P. Urban (Springer, New York, 1972).
[6] B. Dewitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965).
[7] R. Kerner, Ann. Inst. H. Poincaré **9**, 143 (1968).
[8] A. Trautman, Rep. Math. Phys. **1**, 29 (1970).
[9] Y. M. Cho, J. Math. Phys. **16**, 2029 (1975).
[10] A. Salam and J. Strathdee, Ann. Phys. (NY) **141**, 316 (1982).
[11] R. Percacci and S. Randjbar-Daemi, J. Math. Phys. **24**, 807 (1983).
[12] O. M. Moreschi and G. A. J. Sparling, J. Math. Phys. **24**, 303 (1983).
[13] Z. Horvath, L. Palla, E. Cremmer, and J. Scherk, Nucl. Phys. B **127**, 57 (1977); Z. Horvath and L. Palla, *ibid.* **142**, 327 (1978); S. Tanaka, Prog. Theor. Phys. **66**, 1477 (1981); A. Karlhede and T. Tomaras, Phys. Lett. **125**, 49 (1983); S. Randjbar-Daemi, A. Salam, and J. Strathdee, Nucl. Phys. B **214**, 491 (1983); Phys. Lett. B **124**, 349 (1983); C-H. Tze, YTP 83-02 preprint.
[14] R. Arnowitt, S. Deser, and C. W. Misner, Phys. Rev. **122**, 997 (1961).
[15] J. M. Nester, Phys. Lett. A **83**, 241 (1981).
[16] T. Parker and C. H. Taubes, Commun. Math. Phys. **84**, 223 (1982); O. Reula, J. Math. Phys. **23**, 810 (1982).
[17] H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner, Proc. R. Soc. London Ser. A **269**, 21 (1962).
[18] J. Nester and W. Israel, Phys. Lett. A **85**, 259 (1981); M. Ludvigsen and J. Vickers, J. Phys. A **14**, L389 (1981); J. Phys. A **15**, L67 (1982); G. Horowitz and M. Perry, Phys. Rev. Lett. **48**, 31 (1982); O. Reula and K. P. Tod, J. Math. Phys. **25**, 1004 (1984).
[19] A. Ashtekar, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 2.
[20] O. M. Moreschi, Ph.D. thesis, University of Pittsburgh, 1983.
[21] G. W. Gibbons and C. M. Hull, Phys. Lett. B **109**, 190 (1982).
[22] O. M. Moreschi and G. A. J. Sparling, Commun. Math. Phys. **95**, 113 (1984).
[23] L. P. Eisenhart, *Continuous Groups of Transformations* (Dover, New York, 1963).
[24] O. Reula (private communication).
[25] G. W. Gibbons, S. W. Hawking, G. T. Horowitz, and M. J. Perry, Commun. Math. Phys. **88**, 295 (1983).
[26] E. Witten, Commun. Math. Phys. **80**, 381 (1981).
[27] E. Witten, Nucl. Phys. B **195**, 481 (1982).

Conformal quantum Yang–Mills

A. D. Haidari
Department of Physics, University of California, Los Angeles, California 90024

(Received 5 February 1986; accepted for publication 7 May 1986)

Conformally covariant quantization of non-Abelian gauge theory is presented, and the invariant propagators needed for perturbative calculations are found. The vector potential acquires a richer gauge structure displayed in the larger Gupta–Bleuler triplet whose center is occupied by conformal QED. Path integral formulation and BRS invariance are shown on a formal level in one covariant gauge.

I. INTRODUCTION

Conformal invariance of nontrivial QED[1–5] introduces an extra scalar field A_+, which corresponds, in the classical theory, to a Lagrange multiplier needed to fix the gauge in a conformally invariant way.[5] Setting $A_+ = 0$ destroys the covariance of the theory or makes it equivalent to a trivial one that has only longitudinal photons. The five-component vector potential (A_μ, A_+) forms a nondecomposable representation of the conformal group. In the presence of interaction, A_+ couples to an extra current component of degree 4, denoted by J_-. The formal gauge-invariant, minimal coupling introduces unphysical field components in J_-. Taking $J_- = 0$ violates conformal invariance; therefore, we intend to deal with these unphysical fields and quantize them along with the physical ones.

Recently,[5,6] it was shown that these unphysical components are part of a gauge field whose physical subspace is the matter field. This gauge phenomenon reinforces the importance of conformal symmetry in quantum field theory since gauge theories are the most successful ones in describing the interactions of elementary particles. Moreover, the appearance, in a natural way, of these new field components introduces new perturbative diagrams that may have a positive outcome in the renormalization program. Therefore, at this point in the development of the theory, a careful study of renormalization can be very fruitful. Another pressing point is the following: believing in the benefits of the foregoing analysis, which is brought about by requiring conformal invariance of the interaction, then looking at a gauge theory with self-coupling under the same requirements may produce interesting results even on a deeper level. As an example of such a theory, we deal in this paper with conformal Yang–Mills having local SU(N) gauge symmetry. Again, we find that unphysical components of higher degree appear as part of a current that is generated by self-interaction and couples to A_+. We believe that this phenomenon of field components doubling is a general property of conformal charged fields. For example, in Sec. II, it is shown that the conformal electromagnetic potential describes the photon without the need for doubling. However, when the vector potential is charged and self-coupled, new field components surface to enlarge the gauge structure and make it richer than in the Abelian case. The appearance of these new components in the charged fields can be viewed as a recasting of the anomaly in the conformal degree of these fields into higher degree components. It also suggests an alternative to our definition of the unphysical components in which they are multiplied by a weight that depends on the coupling constant and vanishes when the coupling goes to zero. Moreover, the fact that these unphysical fields always couple to A_+ still holds and is very interesting when taken at the level of minimal breaking together with the remarkable property that $\langle 0|A_+(x)\,A_+(x')|0\rangle$ is a constant. In this context, minimal breaking means that the vacuum is not invariant under the action of special conformal transformation but still dilatation invariant (hence, no mass generation is involved). Although it is not shown here, it is believed that the new unphysical components will not contribute to the unitarity of the theory while playing a major role in the renormalization.

In this report, we employ the manifest conformal invariance formalism[7] modified by defining an extension off the Dirac six-cone, which was introduced by us[5] and independently by Ichinose.[8] In Sec. II, we define this extension and show that using this formalism we recover conformal QED. The covariant propagators and the underlying nondecomposable representation of the conformal group are found in Sec. III. The Yang–Mills Lagrangian and nonlinear field equations are written in Sec. IV, where we also outline the path integral formulation and Becchi–Rouet–Stora (BRS) invariance.

The manifest formalism is based on the isomorphism of the conformal group $\mathscr{C}$ to SO(4,2). Therefore, the action of $\mathscr{C}$ in Minkowski space can be linearized by the action of SO(4,2) in R^6, with coordinates $\{y^a\}$, $a = 0,1,2,3,4,5$, and preserving $y^2 = {y_0}^2 - \vec{y}^2 + {y_5}^2$. The two extra dimensions are subsequently eliminated by a constraint ($y^2 = 0$) and a projection ($\lambda y \simeq y$, for $\lambda \neq 0$). The result is the projective Dirac cone—the compactified Minkowski space. Minkowski space is a dense open submanifold whose complement is the light cone at ∞, and with coordinates x^μ defined in the transformation

$$(x^\mu, x^+, x^-) \equiv (y^\mu / y^+, \ln y^+, y^2/(y^+)^2),$$

where $y^\pm = y^5 \pm y^4$. In the coordinates $y_\mu{}^\pm$, the metric

$$(\eta_{\alpha\beta}) = (\eta_{\mu\nu}) \oplus \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix},$$

where $\eta_{\mu\nu} = \mathrm{diag}(+\,-\,-\,-)$ and $\alpha,\beta = 0,1,2,3,+,-$.

If $\{L_{\alpha\beta} = -L_{\beta\alpha}\}$ is a basis for the algebra so(4,2), then the generators of the conformal group are represented by

$$(J_{\mu\nu}, P_\mu, K_\mu, D) \to (L_{\mu\nu}, 2L_{-\mu}, 2L_{+\mu}, 2L_{+-}),$$

where $(J_{\mu\nu}, P_\mu)$ are the Poincaré group generators and D, K_μ are, respectively, the generators of dilatation and conformal boosts.

 0022-2488/86/092409-06$02.50

Minimal weight, K-finite, irreducible representations of SO(4,2) are denoted by $D(E_0,j_1,j_2)$, where E_0 is the "conformal energy" and j_1-j_2 is the helicity. Please see Ref. 1 for details. The conformal scalar and spinor fields carry the following nondecomposable representations (Gupta–Bleuler triplets) of $\mathscr{C}$, respectively[5]:

$$D(3,0,0)\to D(1,0,0)\to D(3,0,0)\,, \tag{1.1}$$

$$D\,(\tfrac{5}{2},\tfrac{1}{2},0,)\to D\,(\tfrac{3}{2},0,\tfrac{1}{2})\to D\,(\tfrac{5}{2},\tfrac{1}{2},0)\,, \tag{1.2}$$

and its helicity conjugate. The arrows denote semidirect sums referred to as "leaks."

II. CONFORMAL QED OFF THE CONE

The incompatibility of conformal invariance with unrestricted gauge invariance in the five-component electrodynamics is evident in the work of many authors.[1–5,9] The scalar A_+ is introduced at the level of gauge fixing, but not before. We intend to deal with this problem in the "extended manifest formalism."

On the projective cone, the electromagnetic action and wave equation are[1]

$$S\,[a,j]=\int (dy)\left[\frac{1}{2}a^\alpha\,\partial^2 a_\alpha - a{\cdot}j\right],$$

$$(dy)=d^6y\,\delta(y^2),\quad \partial^2 a_\alpha = j_\alpha\,. \tag{2.1}$$

In Minkowski notations, they read

$$S\,[A,J\,]=\int d^4x\bigg[\frac{1}{2}A_\mu\Box A^\mu + 2A_-\Box A_+ - 4A_-\,\partial{\cdot}A - 8A_-{}^2 - A_\mu J^\mu - 2A_+J_- - 2A_-J_+\bigg], \tag{2.2}$$

$$\Box A_\mu + 4\,\partial_\mu A_- = J_\mu\,,$$
$$\Box A_+ - 2\,\partial{\cdot}A - 8A_- = J_+\,,$$
$$\Box A_- = J_-\,,$$

where

$$a_\alpha = \frac{\partial x^\beta}{\partial y^\alpha}A_\beta \equiv e^{-x+}\,(V^{-1})_\alpha{}^\beta A_\beta \tag{2.3}$$

and

$$A_\pm = \tfrac{1}{2}\,(A_5 \pm A_4)\,.$$

The free action is invariant under the gauge transformation

$$(A_\mu,A_+,A_-)\to(A_\mu+\partial_\mu\Lambda,A_+,A_- - \tfrac{1}{4}\,\Box\Lambda) \tag{2.4}$$

if Λ is restricted to satisfy $\Box^2\Lambda=0$. Unrestricted gauge invariance of the interaction $(a{\cdot}j)$ gives the following current conservation:

$$\partial_\mu J^\mu + \tfrac{1}{2}\,\Box J_+ = 0\,.$$

The Lorentz condition is

$$y{\cdot}a = A_+ = 0\,,$$

and A_+ is a dipole ghost, $\Box^2 A_+ = 0$.

The solutions of the free wave equation $\partial^2 a_\alpha = 0$ form the followng Gupta–Bleuler triplet (zero-center module[10]):

$$D\,(1,\tfrac{1}{2},\tfrac{1}{2})\to\begin{bmatrix} D(2,1,0)\oplus D(2,0,1)\\ \oplus\, D(0,0,0)\end{bmatrix}\to D\,(1,\tfrac{1}{2},\tfrac{1}{2})\,. \tag{2.5}$$

Note the presence of the identity representation in the physical sector which has no analog in (1.1) or (1.2).

At this point, we attempt to bridge the gap between conformal invariance and unrestricted gauge invariance. We start with a full gauge-invariant, conformally invariant-free action and subsequently introduce gauge fixing that recovers (2.1). We begin by defining an extension for the vector field a_α off the cone in a similar manner to that of the scalar field in Ref. 5. So we modify (2.3) by retaining terms up to order y^2 in the expansion of the fields

$$a_\alpha(y(x))\equiv e^{-x+}\,(V^{-1})_\alpha{}^\beta\,[A_\beta(x_\mu)+x^-B_\beta(x_\mu)]\,. \tag{2.6}$$

The action of special conformal transformation on these fields is

$$K_\mu A_+ = \overset{0}{\nabla}_\mu A_+\,,$$
$$K_\mu A_\nu = \overset{1}{\nabla}_\mu A_\nu + 2(x_\nu A_\mu - \eta_{\mu\nu}x{\cdot}A) + 2\eta_{\mu\nu}A_+\,, \tag{2.7}$$
$$K_\mu A_- = \overset{2}{\nabla}_\mu A_- - A_\mu\,,$$

$$K_\mu B_+ = \overset{2}{\nabla}_\mu B_+ - \partial_\mu A_+\,,$$
$$K_\mu B_\nu = \overset{3}{\nabla}_\mu B_\nu + 2(x_\nu B_\mu - \eta_{\mu\nu}x{\cdot}B) + 2\eta_{\mu\nu}(B_+ - 2A_-) - \partial_\mu A_\nu\,, \tag{2.8}$$
$$K_\mu B_- = \overset{4}{\nabla}_\mu B_- - B_\mu - \partial_\mu A_-\,,$$

where

$$\overset{n}{\nabla}_\mu = x^2\,\partial_\mu - 2x_\mu\,(x{\cdot}\partial + n)\,.$$

Therefore, due to (2.6)–(2.8), the object

$$F_{\alpha\beta} = \partial_\alpha a_\beta - \partial_\beta a_\alpha$$

is a well-defined tensor on the cone—the electromagnetic tensor. Using this tensor, we can construct the following conformally invariant-free action

$$\int (dy)\,(-\tfrac{1}{4}F_{\alpha\beta}F^{\alpha\beta})\,. \tag{2.9}$$

One has to be careful when doing integration by parts, since the integration measure (dy) contains $\delta(y^2)$. Therefore, it is better to write the expressions in Minkowski notation first, then do the manipulation later.

The action (2.9) is invariant under the unrestricted gauge transformation

$$a_\alpha \to a_\alpha + \partial_\alpha\omega\,,$$

where $\omega(y(x)) = \Lambda(x_\mu) + x^-\chi(x_\mu)$. This transformation is equivalent to the following:

$$(A_\mu,A_+,A_-)\to(A_\mu+\partial_\mu\Lambda,A_+,A_-+\chi)\,,$$
$$(B_\mu,B_+,B_-)\to(B_\mu+\partial_\mu\chi,B_+,B_-)\,.$$

The first set is exactly that in (2.4) if we set $\chi = -\tfrac{1}{4}\Box\Lambda$ (i.e., $\partial^2\omega = 0$). Gauge invariance of the interaction gives the following set of conservation laws: $J_+ = 0$, $\partial_\mu J^\mu = 0$. Now we propose to fix the gauge in (2.9) by adding the usual gauge fixing term $-\tfrac{1}{2}(\partial{\cdot}a)^2$ plus some possible invariant piece $\mathscr{Z}$:

$$S_0[a] = \int (dy)\left[-\frac{1}{4}F_{\alpha\beta}F^{\alpha\beta} + \mathscr{L}_{\rm GF}\right],$$

where

$$\mathscr{L}_{\rm GF} = -\tfrac{1}{2}\,(\partial{\cdot}a)^2 + \mathscr{Z}\,. \tag{2.10}$$

In Minkowski notation, this action takes the form

$$S_0[A,B] = \int d^4x[\mathscr{L}_0 + \mathscr{Z} - 2A_+ \partial{\cdot}B - 2B_+(\partial{\cdot}A + 4A_-)]\,,$$

where $\mathscr{L}_0$ is the free Lagrangian in (2.2) with $J_\pm{}^\mu = 0$. Using the conformal transformation (2.7) and (2.8), one can evidently show that the part of the action written explicitly is invariant. So we make the choice

$$\mathscr{Z} = 2A_+ \partial{\cdot}B + 2B_+ \partial{\cdot}A + 8A_-B_+\,, \tag{2.11}$$

in which case we recover the free action in (2.1) or (2.2). So we conclude that in the Abelian case it is possible to write an action with built-in gauge fixing, which is intrinsic on the projective cone, and no auxiliary fields (B_α) are needed for covariant quantization. However, in the non-Abelian theory this is not the case, as we shall see in Sec. IV.

III. THE HOMOGENEOUS PROPAGATORS

The action (2.1) is invariant under the following "gauge" transformation, which we will refer to as the "cone-gauge" (c-gauge, for short):

$$a_\alpha \rightarrow a_\alpha + y^2 b_\alpha\,,$$

where b_α is an arbitrary vector field of degree 3. This is because on the cone ($y^2=0$)

$$\partial^2(a_\alpha + y^2 b_\alpha) = \partial^2 a_\alpha + 4(y{\cdot}\partial + 3)b_\alpha = \partial^2 a_\alpha\,.$$

In fact, the solutions (2.5) of the free wave equation form the quotient of the nondecomposable representation of $\mathscr{C}$

$$\{\text{CQED}\} \rightarrow \{D(3,\tfrac{1}{2},\tfrac{1}{2}) \rightarrow D(4,0,0)\}\,,$$

where {CQED} is the module (2.5) and the invariant "c-gauge" subspaces are all of the form $y^2 b_\alpha$. To find the extension to the full triplet, we investigate the free propagator $K_{\alpha\beta}(y{\cdot}y') = \langle a_\alpha(y)\, a_\beta(y')\rangle$, where the $a_\alpha(y)$ are the quantum field operators. An immediate candidate for this propagator is suggested by comparing the free wave equation with that of the scalar ($\partial^2\Phi = 0$). That is, we set $K_{\alpha\beta} = -\eta_{\alpha\beta}K$, where K is the scalar propagator[5]:

$$K = (y{\cdot}y')^{-1} + \lambda y^2 y'^2 (y{\cdot}y')^{-3}\,,$$

and λ is a dimensionless real parameter. Then, $K_{\alpha\beta}$ carries the representation

$$D_6 \otimes [D(3,0,0) \rightarrow D(1,0,0) \rightarrow D(3,0,0)]\,, \tag{3.1}$$

where D_6 is the finite-dimensional vector representation of SO(4,2).

However, this propagator is not "clean," in the sense that it contains "spectator ghosts" carrying representations that are not Weyl-equivalent[1] to the physical ones and appear as direct sum representations.

Proof: Let $\widehat{C}_{\alpha\beta}$ be the second-order Casimir operator in the vector representation. Then $\widehat{C}_{\alpha\beta} - c\eta_{\alpha\beta} = 0$ in an irreducible representation for some constant c. However, if the representation is nondecomposable, then $(\widehat{C} - c)$ is nilpotent and only $(\widehat{C} - c)^n = 0$ for some positive integer n less than or at most equal to the number of levels of leak in the representation. One can easily check that

$$[(\widehat{C} - c)^n]_\alpha{}^\beta K_{\beta\gamma} = [(\widehat{C} - c)^n]_{\alpha\gamma} K \neq 0, \quad \forall\{n,c\}\,.$$

In our search for the free propagator, we require that it satisfies the following conditions: (i) SO(4,2) invariant; (ii) homogeneous of degree 1 and linear in y^2; (iii) contains the physical and Weyl-equivalent representations; and (iv) satisfies the Casimir equation $[(\widehat{C} - c)^n]_\alpha{}^\beta K^{(n)}_{\beta\gamma} = 0$. These requirements fix $K_{\alpha\beta}$ uniquely modulo c-gauge. The third one implies that $c = 0$, since the physical representation $D(2,1,0) \oplus D(2,0,1)$ is Weyl equivalent to $D(0,0,0)$ as seen in (2.5). The following propagators are solutions to the Casimir equation:

$$K^{(1)}_{\alpha\beta} = y^2 y'^2 y_\alpha y'_\beta (y{\cdot}y')^{-4}\,,$$

$$K^{(2)}_{\alpha\beta} = y^2 y'^2 [\eta_{\alpha\beta}(y{\cdot}y')^{-3} - y'_\alpha y_\beta (y{\cdot}y')^{-4} + \xi y_\alpha y'_\beta (y{\cdot}y')^{-4}]\,, \tag{3.2}$$

$$K^{(n)}_{\alpha\beta} = \eta_{\alpha\beta}(y{\cdot}y')^{-1} + \tfrac{1}{3}(y_\alpha y'_\beta - y'_\alpha y_\beta)(y{\cdot}y')^{-2} + y^2 y'^2 \{\lambda \eta_{\alpha\beta}(y{\cdot}y')^{-3} + [\xi y_\alpha y'_\beta - (\lambda + \tfrac{1}{3})\, y'_\alpha y_\beta]\,(y{\cdot}y')^{-4}\}, \quad n \geqslant 3\,, \tag{3.3}$$

where λ and ξ are arbitrary constants whose values can be chosen later to eliminate the most singular distributions.

It is evident that $K^{(2)}_{\alpha\beta}$ is the "c-gauge" propagator, since it is made up of modes all of the form $y^2 b_\alpha$. The nontrivial propagator (which contains a nonvanishing transverse part) is $K^{(n)}_{\alpha\beta}$ in (3.3). Therefore, the free propagator is determined, up to a "null propagator," by taking $K_{\alpha\beta} = K^{(3)}_{\alpha\beta}$.

It is interesting to note that restricting (3.3) to the cone, we recover the propagator for "gradient-type" gauge theory found by Binegar *et al.*[1] in describing conformal QED, and denoted by $K^{q-}_{\alpha\beta}$. Here, $q = \tfrac{2}{3}$, which is just the right value needed to remove one of the two ghosts in (3.1). The propagator (3.2) without $y^2 y'^2$ factor is what they call the propagator for "current-type" gauge theory, $K^{q+}_{\alpha\beta}$ (here $q = \xi$), which carries the triplet

$$D(4,0,0) \rightarrow D(3,\tfrac{1}{2},\tfrac{1}{2}) \rightarrow D(4,0,0)\,.$$

Now $K_{\alpha\beta}$ satisfies the following equations (mod y^2):

$$\partial^\alpha K_{\alpha\beta} = 0\,, \tag{3.4}$$

$$y^\alpha \partial^2 K_{\alpha\beta} = 0\,. \tag{3.5}$$

These conditions actually eliminate the "junk" in (3.1). Analysis of the modes in the Fourier expansion of $K_{\alpha\beta}$ shows that the triplet (2.5) forms the center of a larger Gupta–Bleuler triplet:

$$\begin{array}{c} D(4,0,0) \\ \downarrow \\ D(3,\tfrac{1}{2},\tfrac{1}{2}) \\ \downarrow \\ \boxed{D(1,\tfrac{1}{2},\tfrac{1}{2}) \rightarrow \left\{ \begin{array}{c} D(2,1,0) \oplus D(2,0,1) \\ \oplus\ \text{Id} \end{array} \right\} \rightarrow D(1,\tfrac{1}{2},\tfrac{1}{2})} \\ \downarrow \\ D(3,\tfrac{1}{2},\tfrac{1}{2}) \\ \downarrow \\ D(4,0,0) \end{array}\,, \tag{3.6}$$

where the representation A_α is (2.6) carries the top $\tfrac{2}{3}$ of the triplet. The free propagator (3.3) is equivalent to the following set of nonvanishing two-point functions $(\lambda+\frac{1}{3}=\xi=-\frac{1}{2})$:

$$\langle A_\mu(x)\,A_\nu(x')\rangle = r^{-2}\,(-\eta_{\mu\nu}+\tfrac{2}{3}r_\mu r_\nu r^{-2})\,,$$
$$\langle A_\mu(x)\,A_+(x')\rangle = \tfrac{2}{3}r_\mu r^{-2},\quad \langle A_\mu(x)\,A_-(x')\rangle = \tfrac{1}{3}r_\mu r^{-4}\,,$$
$$\langle A_+(x)\,A_-(x')\rangle = -\tfrac{1}{3}r^{-2},\quad \langle A_+(x)\,A_+(x')\rangle = \tfrac{1}{3}\,, \tag{3.7}$$
$$\langle B_\mu(x)\,B_\nu(x')\rangle = \tfrac{4}{3}\eta_{\mu\nu}\,r^{-6},\quad \langle B_+(x)\,B_-(x')\rangle = \tfrac{2}{3}r^{-6}\,, \tag{3.8}$$
$$\begin{aligned}&\langle A_\mu(x)\,B_\nu(x')\rangle = r^{-4}\,(-\eta_{\mu\nu}+\tfrac{4}{3}r_\mu r_\nu r^{-2})\,,\\ &\langle A_\mu(x)\,B_+(x')\rangle = \langle B_\mu(x)\,A_+(x')\rangle = \tfrac{2}{3}r_\mu r^{-4}\,,\\ &\langle A_\mu(x)\,B_-(x')\rangle = \langle B_\mu(x)\,A_-(x')\rangle = \tfrac{2}{3}r_\mu r^{-6}\,,\\ &\langle A_+(x)\,B_-(x')\rangle = -\tfrac{1}{3}r^{-4}\,,\end{aligned} \tag{3.9}$$

where

$$r_\mu = x_\mu - x'_\mu\,.$$

IV. CONFORMAL YANG–MILLS

Conformal Yang–Mills was considered by Zaikov,[11] who also found it necessary to introduce auxiliary fields. Fradkin and Palchik[12] found a nonlinear, nonlocal transformation of the fields. In this paper, we deal with conformal Yang–Mills in the extended manifest formalism.

The problem here is the same as the one in conformal scalar and spinor QED. The free Lagrangian is "intrinsic" on the cone, which means that the unphysical fields do not contribute, but the interaction is not.

The general form of the pure Yang–Mills Lagrangian is

$$\mathscr{L} = \mathscr{L}_0 + g\mathscr{L}_1 + g^2\mathscr{L}_2\,,$$

where g is a dimensionless coupling constant, $\mathscr{L}_0$ is quadratic, $\mathscr{L}_1$ cubic, and $\mathscr{L}_2$ quartic in the field. The $\mathscr{L}_0$ and $\mathscr{L}_2$ can be written intrinsically on the cone, but not $\mathscr{L}_1$. The general form of $\mathscr{L}_1$ is $(a^\alpha a^\beta D_\alpha a_\beta)$, where D_α is a linear differential operator of degree 1. The only such operator intrinsic on the cone is $D_\alpha = y_\alpha\,\partial^2 - 2(y\cdot\partial+2)\,\partial_\alpha$, which is not satisfactory since it does not reproduce the three-gluon vertex, while $D_\alpha=\partial_\alpha$ does. Therefore, an extension off the cone is needed and the unphysical field B_α will contribute. So we adopt (2.6)–(2.8), where as usual the vector potential a_α is in the adjoint representation of SU(N):

$$a_\alpha(y) = a_\alpha{}^i(y)T_i,\quad i=1,\ldots,(N^2-1)\,.$$

The $\{T_i\}$ are the Hermitian generators of SU(N) with a Lie algebra and normalization

$$[T_i,T_j] = iC_{ij}{}^kT_k,\quad \mathrm{Tr}(T_iT_j) = \tfrac{1}{2}\delta_{ij}\,.$$

Define the covariant derivative and its commutator

$$\mathscr{D}_\alpha = \mathbb{1}\,\partial_\alpha + iga_\alpha\,,$$
$$F_{\alpha\beta}\equiv(1/ig)\,[\mathscr{D}_\alpha,\mathscr{D}_\beta] = \partial_\alpha a_\beta - \partial_\beta a_\alpha + ig[a_\alpha,a_\beta]\,,$$

where $\mathbb{1}$ is the unit matrix.

The pure Yang–Mills invariant action is

$$\begin{aligned}S[a] &= \int(dy)\,\mathrm{Tr}\Big[-\frac{1}{4}F_{\alpha\beta}F^{\alpha\beta}\Big]\\ &= \int(dy)\,\mathrm{Tr}\Big\{-\frac{1}{2}\partial_\alpha a_\beta(\partial^\alpha a^\beta - \partial^\beta a^\alpha)\\ &\qquad - ig[a^\alpha,a^\beta]\,\partial_\alpha a_\beta + \frac{1}{4}g^2[a_\alpha,a_\beta]\,[a^\alpha,a^\beta]\Big\}\,,\end{aligned} \tag{4.1}$$

$$S[A,B] = \int d^4x\,[(\mathscr{L}_0-\mathscr{L}_{\mathrm{GF}}) + g\mathscr{L}_1 + g^2\mathscr{L}_2]\,, \tag{4.2}$$

where $\mathscr{L}_{\mathrm{GF}}$ is the trace of (2.10) and (2.11), and in the free theory $\mathscr{L}_{\mathrm{GF}} = \tilde{\mathscr{L}}$ since $\partial\cdot a=0$ as shown in (3.4):

$$\begin{aligned}\mathscr{L}_0 &= \mathrm{Tr}(\tfrac{1}{2}a^\alpha\,\partial^2 a_\alpha)\\ &= \mathrm{Tr}(\tfrac{1}{2}A_\mu\Box A^\mu + 2A_-\Box A_+ - 4A_-\,\partial\cdot A - 8A_-{}^2),\\ \mathscr{L}_1 &= -i\,\mathrm{Tr}\{[A^\mu,A^\nu]\,\partial_\mu A_\nu + 2[A^\mu,A_+]\,\partial_\mu A_-\\ &\quad + 2[A^\mu,A_-]\,\partial_\mu A_+ + 2A_+[A_\mu,B^\mu]\\ &\quad + 4A_+[A_-,B_+]\}\,,\\ \mathscr{L}_2 &= \mathrm{Tr}\{\tfrac{1}{4}\,[A_\mu,A_\nu]^2 - 2[A_+,A_-]^2\\ &\quad + 2[A_\mu,A_+]\,[A^\mu,A_-]\}\,.\end{aligned}$$

Each term in (4.2) is separately invariant. Note, however, that the unphysical fields B_α that appear in $\mathscr{L}_1$ cannot be extracted into an invariant piece as in the Abelian case.

The action (4.1) is invariant under the gauge transformation

$$a_\alpha \to \Omega a_\alpha\Omega^{-1} - (i/g)\Omega(\partial_\alpha\Omega^{-1})\,, \tag{4.3}$$

where

$$\Omega = \exp(-ig\omega),\quad \omega(y(x)) = [\Lambda^i(x_\mu) + x^-\chi^i(x_\mu)]T_i\,.$$

The infinitesimal form of (4.3) is

$$\delta a_\alpha = \partial_\alpha\omega + ig[a_\alpha,\omega] = [\mathscr{D}_\alpha,\omega]\,.$$

In the Abelian case, there is no self-coupling and the free field, by virture of Eq. (3.4), satisfies $\partial\cdot a=0$. Therefore, in the gauge transformation $\delta a_\alpha = \partial_\alpha\omega$, this requires $\partial^2\omega=0$, which reads $\chi = -\frac{1}{4}\Box\Lambda$; in agreement with the choice made in Sec. II for QED.

The infinitesimal gauge transformation amounts to the following:

$$\begin{aligned}&\delta A_\mu = \partial_\mu\Lambda + ig[A_\mu,\Lambda] = [D_\mu,\Lambda]\,,\\ &\delta A_+ = ig[A_+,\Lambda],\quad \delta A_- = \chi + ig[A_-,\Lambda]\,,\\ &\delta B_\mu = ig[B_\mu,\Lambda] + [D_\mu,\chi]\,,\\ &\delta B_\pm = ig\{[B_\pm,\Lambda] + [A_\pm,\chi]\}\,,\end{aligned}$$

and

$$D_\mu = \mathbb{1}\,\partial_\mu + igA_\mu\,.$$

We fix the gauge in (4.1) and (4.2) by adding $\mathscr{L}_{\mathrm{GF}}$, and get the following conformally covariant nonlinear equations upon variations of A_α:

$$\begin{aligned}&[D^\nu,f_{\nu\mu}] + \partial_\mu\,\partial\cdot A + 4\,\partial_\mu A_- - 2ig\{[A_+,[D_\mu,A_-]]\\ &\quad + [A_-,[D_\mu,A_+]] + [B_\mu,A_+]\} = J_\mu\,,\end{aligned} \tag{4.4a}$$
$$\begin{aligned}&[D^\mu,[D_\mu,A_-]] - ig\{[A_\mu,B^\mu] + 2[A_-,B_+]\}\\ &\quad + 2g^2[A_-,[A_-,A_+]] = J_-\,,\end{aligned} \tag{4.4b}$$
$$\begin{aligned}&[D^\mu,[D_\mu,A_+]] - 2\,\partial\cdot A - 8A_- + 2ig[A_+,B_+]\\ &\quad + 2g^2[A_+,[A_+,A_-]] = J_+\,,\end{aligned} \tag{4.4c}$$

where

$$f_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + ig[A_\mu, A_\nu],$$

and sources have been included. Variations with respect to B_α give the equation $gA_+ = 0$, which is not an equation for the free theory and can only be imposed as initial conditions on the physical subspace.

The propagators for the free quantum fields are just those given in (3.7)–(3.9) with superscripts i, j, and δ^{ij} multiplying the right side. The last set (3.9) is incompatible with two of the free wave equations obtained from (4.4a) and (4.4b) with $g = 0$. The reason is as was stated below (3.6) in that the A_α carry the "physical" modes that do satisfy the free wave equations, but they also carry the "scalar" modes that do not. To cure this, one may follow one of two procedures presented in Ref. 5. The first makes use of dimensional regularization and the second introduces logarithmic modes in the "scalar" sector. However, we will not pursue this any further in this report since it does not pose any problem to the physics and it only amounts to a better choice of field variables that splits A_α into its two components, the "physical" and "scalar."

In the presence of matter, gauge invariance of the interaction $\mathrm{Tr}(a \cdot j)$ gives the following set of conservation laws for the current:

$$J_+ = 0 \quad \text{and} \quad [D_\mu, J^\mu] + 2ig[A_+, J_-] = 0. \qquad (4.5)$$

The spinor action and six-current are[5]

$$S[\psi, \sigma] = \int d^4x \left[-\frac{i}{2} \bar{\psi} \overleftrightarrow{\partial} \psi + g\bar{\psi} A \psi + g(\bar{\psi} A_+ \sigma + \bar{\sigma} A_+ \psi) \right],$$

$$J_+{}^i = 0, \quad J_\mu{}^i = -g\bar{\psi}\gamma_\mu T^i \psi,$$

$$J_-{}^i = -(g/2)(\bar{\psi} T^i \sigma + \bar{\sigma} T^i \psi),$$

and (4.5) is satisfied by virture of the field equations. The action of special conformal transformation on these spinors is

$$K_\mu \psi = \overset{3/2}{\nabla}{}_\mu \psi - \tfrac{1}{2}[\gamma_\mu, x \cdot \gamma]\psi,$$

$$K_\mu \sigma = \overset{5/2}{\nabla}{}_\mu \sigma - \tfrac{1}{2}[\gamma_\mu, x \cdot \gamma]\sigma - \gamma_\mu \psi.$$

The homogeneous two-point functions are

$$\langle \psi(x)\, \bar{\psi}(x') \rangle = ir \cdot \gamma\, r^{-4},$$

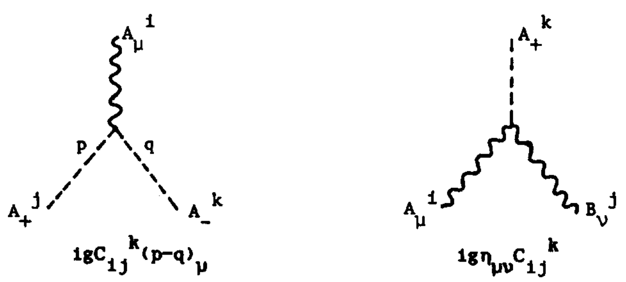
$$\langle \psi(x)\, \bar{\sigma}(x') \rangle = i\, r^{-4},$$

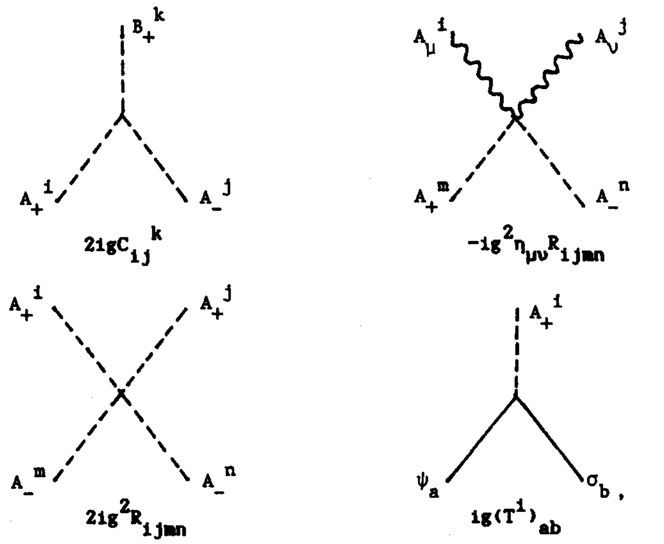
$$\langle \sigma(x)\, \bar{\sigma}(x') \rangle = 0.$$

The total Yang–Mills Lagrangian is

$$\mathscr{L}_{\mathrm{YM}} = \mathscr{L}_0 + g\mathscr{L}_1 + g^2 \mathscr{L}_2 - (i/2)\bar{\psi} \overleftrightarrow{D} \psi + g(\bar{\psi} A_+ \sigma + \bar{\sigma} A_+ \psi).$$

The new vertices are





where

$$R_{ijmn} = C_{im}{}^k C_{jnk} + C_{in}{}^k C_{jmk}$$

and $a, b = 1, \ldots, N$.

Notice that the unphysical fields B_α and σ always couple to A_+. It remains to be shown that the coupling of the unphysical fields do not contribute to the physical processes.

Finally, we outline the path integral formulation in one gauge using the Faddeev–Popov ansatz and indicate the residual BRS invariance.

The generating functional for the Green's functions is

$$Z[j] = \mathscr{N} \int (\mathscr{D}a_\alpha) \exp i \left\{ \int (dy) \mathrm{Tr} \left[-\frac{1}{4} F_{\alpha\beta} F^{\alpha\beta} - a \cdot j \right] \right\};$$

then insert $1 = \int (\mathscr{D}\omega)\, \Delta_{\mathrm{FP}}[a_\alpha{}^\omega]\, \delta(\partial \cdot a^\omega - \phi)$, which defines the Faddeev–Popov determinant Δ_{FP}.

Gauge invariance leaves a meaningless infinite volume integral in gauge space that is absorbed in the normalization factor $\mathscr{N}$. Then, the constant

$$\int (\mathscr{D}\phi) \exp\left[-\frac{i}{2} \int (dy)\ \mathrm{Tr}\, \phi^2 \right]$$

is inserted and integration over ϕ is carried out to result in the following amplitude:

$$Z[j] = \mathscr{N}' \int (\mathscr{D}a_\alpha)\, \Delta_{\mathrm{FP}}[a] \times \exp i \left\{ \int (dy) \mathrm{Tr} \left[-\frac{1}{4} F_{\alpha\beta} F^{\alpha\beta} - \frac{1}{2}(\partial \cdot a)^2 - a \cdot j \right] \right\}.$$

Hence $\Delta_{\mathrm{FP}}[a] = \det\{M[a]\}$, where $M[a]$ is the operator

$$\delta(y^2) M(y, y') = \delta^6(y - y')\, (\partial^\alpha \mathscr{D}_\alpha).$$

Then, we express the determinant as the integral

$$\Delta_{\mathrm{FP}}[a] = \int (\mathscr{D}\eta_i)\, (\mathscr{D}\eta_i{}^\dagger) \times \exp i \left\{ \int (dy)\ \mathrm{Tr}(\eta^\dagger\, \partial^\alpha [\mathscr{D}_\alpha, \eta]) \right\}$$

over $2(N^2-1)$ Grassmann scalars η_i and $\eta_i{}^\dagger$, referred to as FP ghosts. Finally, we obtain the generating functional

$$Z[j,k,k^\dagger]=\mathscr{N}\int(\mathscr{D}a\mathscr{D}\eta\mathscr{D}\eta^\dagger)$$

$$\times\exp iS[a,\eta,\eta^\dagger,j,k,k^\dagger],$$

$$S=\int(dy)\,\mathrm{Tr}\Big\{-\frac{1}{4}F_{\alpha\beta}F^{\alpha\beta}-\frac{1}{2}(\partial\cdot a)^2+\eta^\dagger\,\partial^\alpha[\mathscr{D}_\alpha,\eta]$$

$$-a\cdot j-\eta^\dagger K-\eta K^\dagger\Big\}, \tag{4.6}$$

where we have included sources for the ghosts.

The effective action (4.6) is not invariant under the gauge transformation (4.3), but it is invariant under the rigid BRS transformations

$$\delta a_\alpha=\theta[\mathscr{D}_\alpha,\eta],$$

$$\delta\eta=(g/2i)\theta[\eta,\eta],$$

$$\delta\eta^\dagger=\theta(\partial\cdot a),$$

and θ is a Grassmann constant parameter. This invariance is used to derive the Ward identities, which may differ from Slavnov–Taylor identities for ordinary Yang–Mills. But this question is beyond the scope of this paper. We also leave, for a possible future study, the question of choosing a better gauge fixing term (e.g., $\mathscr{L}_{\mathrm{GF}}$) and the resulting BRS invariance. Therefore, no attempt is made here to write (4.6) in Minkowski notation and further the investigation.

ACKNOWLEDGMENTS

It is a pleasure to thank C. Fronsdal for guidance in this work and very helpful discussions. Fruitful conversations with M. Flato are greatly appreciated. Discussions with B. Binegar are deeply acknowledged.

The author thanks the Saudi Ministry of Higher Education and the University of Petroleum and Minerals for support.

[1]B. Binegar, C. Fronsdal, and W. Heindenreich, J. Math. Phys. **24**, 2828 (1983).

[2]R. Zaikov, JINR Preprint E2-83-28, Dubna, 1983.

[3]P. Furlan, V. Petkova, G. Sotkov, and I. Todorov, ISAS Preprint 51/84/EP, Trieste, 1984.

[4]V. Petkova, G. Sotkov, and I. Todorov, Commun. Math. Phys. **97**, 227 (1985).

[5]F. Bayen, M. Flato, C. Fronsdal, and A. Haidari, Phys. Rev. D **32**, 2673 (1985).

[6]W. Heidenreich, Nuovo Cimento A **80**, 220 (1984).

[7]P. A. M. Dirac, Ann. Math. **37**, 429 (1936); G. Mack and A. Salam, Ann. Phys. (NY) **53**, 174 (1969).

[8]S. Ichinose, RIMS preprint 503, Kyoto, 1985.

[9]G. Sotkov and D. Stoyanov, J. Phys. A **16**, 2817 (1983).

[10]*Essays on Supersymmetry*, edited by C. Fronsdal (Reidel, Dordrecht, 1986), p. 123.

[11]R. Zaikov, JINR Preprint E2-83-44, Dubna 1983.

[12]E. Fradkin and M. Palchik, Phys. Lett. B **147**, 86 (1984).

Gauge theory of a group of diffeomorphisms. I. General principles

Eric A. Lord
Department of Applied Mathematics, Indian Institute of Science, Bangalore 560012, India
P. Goswami
Department of Physics, Indian Institute of Science, Bangalore 560012, India

(Received 25 July 1985; accepted for publication 30 April 1986)

Any $(N+M)$-parameter Lie group G with an N-parameter subgroup H can be realized as a global group of diffeomorphisms on an M-dimensional base space B, with representations in terms of transformation laws of fields on B belonging to linear representations of H. The gauged generalization of the global diffeomorphisms consists of general diffeomorphisms (or coordinate transformations) on a base space together with a local action of H on the fields. The particular applications of the scheme to space-time symmetries is discussed in terms of Lagrangians, field equations, currents, and source identities.

I. INTRODUCTION

The theory of Yang and Mills[1] provides a prescription for "gauging" an internal symmetry group. The linear action of the group on physical fields is generalized from a "global" action to a "local" action by the introduction of auxiliary fields—the so-called Yang–Mills potentials or connection coefficients for the group in question. This idea was applied to the group of Lorentz rotations of an orthonormal tetrad (in a metric space-time) by Utiyama[2] and Sciama.[3] It is natural to regard the connection coefficients in this case as the anholonomic components of the linear connection of the space-time. The holonomic linear connection is then metric compatible and asymmetric (the space-time is a U_4, in Hehl's terminology[4]).

The Poincaré group (group of isometries of Minkowski space-time) lies outside the scope of the original Yang–Mills theory because it acts on the space-time as well as on physical fields. Nevertheless, as was shown by Kibble,[5] it can be gauged. The auxiliary fields consist of a tetrad and a connection for the Lorentz rotations of the tetrad. The action of the gauged Poincaré group is the action of general coordinate transformations (or, interpreted actively, space-time diffeomorphisms) together with Lorentz rotations of the tetrad. It is natural then to define the space-time metric to be the one with respect to which the tetrad is orthonormal and to define the linear connection of the space-time to be the one arising from the Lorentz connection. We obtain again the U_4 theory of Utiyama and Sciama. However, the tetrad and the general coordinate transformations arise out of the gauge principle in Kibble's approach; moreover, the metric (and not just the linear connection) is constructed from the auxiliary fields—in the Lorentz gauge theory of Utiyama and Sciama, the tetrad, the metric, and the general coordinate transformations were presupposed *ab initio* and were extraneous to the gauge principle.

The effect of an infinitesimal space-time diffeomorphism (or general coordinate transformation) on an anholonomic field (i.e., a set of scalars that transform linearly and homogeneously under tetrad rotations) is the same as the effect of an infinitesimal *parallel transport* combined with an infinitesimal tetrad rotation. The work of von der Heyde[6] and the subsequent developments of Poincaré gauge theory by Hehl and co-workers[4] have revealed that it is this parallel transport action, rather than space-time diffeomorphisms or coordinate transformations, that should be regarded as the translational part of the gauged Poincaré action. With this interpretation, the tetrad is itself a set of Yang–Mills potentials, constituting the connection for the translational subgroup. The Yang–Mills "field strengths" are the torsion (translational part) and the curvature (rotational part) of the U_4.

Kibble's approach can be applied to more general groups. Some general aspects of the gauging of space-time diffeomorphisms have been worked out by Harnad and Pettitt.[7]

As was shown by Lord,[8] the gauge theory of the affine group, together with a space-time metric imposed as an extraneous field, is equivalent to a purely holonomic metric-affine theory.[9] Of course, the affine extension of the action of the Poincaré group on Minkowski space-time cannot be a symmetry group for a physical theory, but that does not rule out the possibility of the existence of gauge potentials for the group. Indeed, there are some indications that the affine extension of Poincaré gauge theory may be the correct extension required for an understanding of the relationship between strong and gravitational interactions.[10] Poincaré gauge theory has been treated as a limiting case of a de Sitter gauge theory in the work of MacDowell and Mansouri.[11] The gauging of the conformal group (group of diffeomorphisms of Minkowski space-time that preserve the light-cone structure) is usually discussed in the language of fiber bundles, employing second-order frames.[12] As will become clear from the present work, the concept of second-order frames is by no means essential in a gauge theory of the conformal group.

Apart from the extensions mentioned above, there is also the interesting possibility of extending the Poincaré gauge theory so as to include internal symmetries in a nontrivial way.[13]

For a more exhaustive survey of the literature on Poincaré gauge theory and its extensions, the reader is referred to

 0022-2488/86/092415-08$02.50

the review article of Ivanenko and Sardanashvily.[14]

We shall present a general geometrical framework that includes the above-mentioned theories as particular cases. The central idea is the following: Let G be an $(N+M)$-parameter Lie group possessing an N-parameter subgroup H. Let ψ be a set of fields on an M-dimensional space B, belonging to a linear representation of H. Then G can be realized as a global group of diffeomorphisms on B together with a representation in terms of transformation laws for ψ. Moreover, this can be done in such a way that the gauging of the global diffeomorphism group leads to a local transformation group consisting of general diffeomorphisms on B together with an intrinsic action of H on ψ.

Some aspects of our formalism have been foreshadowed in the work of Harnad and Pettitt.[7] However, the present work goes beyond the scheme of Harnad and Pettitt in several respects, and our approach is different. We do not begin with a global group of diffeomorphisms and attack the problem of gauging it—we begin with a full-fledged gauge theory and come to a global diffeomorphism group as a limiting case.

II. STRUCTURE OF THE GROUP AND ITS POTENTIALS

Consider an $(N+M)$-parameter Lie group G with generators π_α, G_a satisfying the commutation relations

$$\begin{aligned}
&[\pi_\alpha,\pi_\beta] = c_{\alpha\beta}{}^{\gamma}\pi_\gamma + c_{\alpha\beta}{}^{c}G_c,\\
&[\pi_\alpha,G_b] = c_{ab}{}^{\gamma}\pi_\gamma + c_{ab}{}^{c}G_c, \qquad (2.1)\\
&[G_a,G_b] = c_{ab}{}^{c}G_c.
\end{aligned}$$

The Greek indices $\alpha,\beta,\ldots$ are M-fold indices and the Latin indices $a,b,\ldots$ are N-fold. The N-parameter subgroup generated by the G_a will be called H. We shall also employ $(N+M)$-fold indices $A,B,\ldots$, in terms of which (2.1) is

$$[G_A,G_B] = c_{AB}{}^{C}G_C \qquad (2.2)$$

(where $\pi_\alpha = G_\alpha$, $c_{ab}{}^{\gamma} = 0$).

We shall set up a gauge theory of the group G on an M-dimensional base space. We shall take $M=4$ with a view to the physical applications in which the base space is space-time. However, it should be borne in mind that the geometrical framework is valid for any M and thus has potentially wider applications.

To begin with, we regard G as a group that acts on fields over space-time but not on space-time points. The infinitesimal action of G on a field Ψ belonging to a linear representation of G will be written

$$\delta\Psi = \epsilon\Psi, \qquad (2.3)$$

where

$$\epsilon = \epsilon^A G_A = \epsilon^\alpha\pi_\alpha + \epsilon^a G_a. \qquad (2.4)$$

Here, π_α and G_a denote the matrix representatives of the corresponding generators.

The group G is gauged in the standard Yang–Mills way be introducing Yang–Mills potentials, which are the coefficients of a connection

$$\Gamma_i = \Gamma_i{}^A G_A = e_i{}^\alpha\pi_\alpha + \Gamma_i{}^a G_a. \qquad (2.5)$$

The Latin letters $i,j,\ldots$ will be used for holonomic space-time indices. The covariant derivative of a field Ψ,

$$\nabla_i\Psi = \partial_i\Psi - \Gamma_i\Psi, \qquad (2.6)$$

transforms like Ψ under the action of an element of G with space-time-dependent parameters provided the connection has the transformation law

$$\delta\Gamma_i = \nabla_i\epsilon = \partial_i\epsilon - [\Gamma_i,\epsilon], \qquad (2.7)$$

which corresponds to the transformation law

$$\delta\Gamma_i{}^A = \nabla_i\epsilon^A = \partial_i\epsilon^A + \epsilon^B\Gamma_i{}^C c_{BC}{}^A, \qquad (2.8)$$

for the Yang–Mills potentials. The Yang–Mills field strengths are the coefficients of the curvature

$$\begin{aligned}
G_{ij} &= \partial_i\Gamma_j - \partial_j\Gamma_i - [\Gamma_i,\Gamma_j]\\
&= G_{ij}{}^A G_A = G_{ij}{}^\alpha\pi_\alpha + G_{ij}{}^a G_a, \qquad (2.9)
\end{aligned}$$

which leads to

$$G_{ij}{}^A = \partial_i\Gamma_j{}^A - \partial_j\Gamma_i{}^A - \Gamma_i{}^B\Gamma_j{}^C C_{BC}{}^A. \qquad (2.10)$$

The fields strengths have a linear homogeneous transformation law

$$\delta G_{ij} = [\epsilon,G_{ij}], \quad \delta G_{ij}{}^A = \epsilon^B G_{ij}{}^C c_{BC}{}^A \qquad (2.11)$$

and satisfy the Bianchi identities

$$\nabla_{[i}G_{jk]}{}^A = 0. \qquad (2.12)$$

Now let ψ be a field belonging to a linear representation R of the subgroup H. We write the infinitesimal transformation law of ψ under the action of H in the form

$$\delta\psi = \epsilon^a\overline{G}_a\psi, \qquad (2.13)$$

where the $\overline{G}_a$ are the matrix representatives of the generators, in the representation R. *In general, it is not possible to extend R to a representation of G.* Three particular representations of H, deducible from the structure constants of G, are an $(N+4)$-dimensional representation T, a four-dimensional representation S, and an N-dimensional representation C (the adjoint representation of H), generated, respectively, by the matrices T_a, S_a, and C_a defined by

$$(T_a)_B{}^C = C_{Ba}{}^C, \quad (S_a)_\beta{}^\gamma = c_{\beta a}{}^\gamma, \quad (C_a)_b{}^c = c_{ba}{}^c. \qquad (2.14)$$

(The relations $[T_a,T_b] = c_{ab}{}^cT_c$, $[S_a,S_b] = c_{ab}{}^cS_c$, and $[C_a,C_b] = c_{ab}{}^cC_c$ are consequences of the Jacobi identities for the generators of G.)

Observe that if

$$c_{ab}{}^c = 0, \qquad (2.15)$$

the representation T is just the direct sum of the representations S and C. The relation (2.15) holds, for instance, when G is the Poincaré group, the affine group, or the de Sitter group. If (2.15) does not hold, then T is reducible but not completely reducible. An example of this is $G = SO(4,2)$, a circumstance that leads to interesting special features for the gauge theory of the conformal group.

The infinitesimal transformation laws for fields belonging to the representations T, S, and C of H, under the action of H, are, respectively,

$$\begin{aligned}
&\delta\phi_A = \epsilon^b c_{Ab}{}^C\phi_C, \quad \delta\chi_\alpha = \epsilon^b c_{ab}{}^\gamma\chi_\gamma,\\
&\delta\psi_a = \epsilon^b c_{ab}{}^c\psi_c. \qquad (2.16)
\end{aligned}$$

The contragredient representations $\widetilde{T}$, $\widetilde{S}$, and $\widetilde{C}$ have the corresponding infinitesimal transformation laws

$$\delta\phi^A = -\epsilon^b\phi^C c_{Cb}{}^A,\ \delta\chi^\alpha = -\epsilon^b\chi^\gamma c_{\gamma b}{}^\alpha, \qquad (2.17)$$
$$\delta\psi^a = -\epsilon^b\psi^c c_{cb}{}^a.$$

Observe in particular that, if χ_α transforms according to S, we can extend it to a χ_A transforming according to T, simply by defining $\chi_a = 0$.

Under the action of H, the Yang–Mills potentials of G have the transformation laws

$$\delta e_i{}^\alpha = -\epsilon^b e_i{}^\gamma c_{\gamma b}{}^\alpha, \qquad (2.18)$$

$$\delta\Gamma_i{}^a = \partial_i\epsilon^a + \epsilon^b\Gamma_i{}^c c_{bc}{}^a - \epsilon^b e_i{}^\gamma c_{\gamma b}{}^a. \qquad (2.19)$$

In regions of space-time where the 4×4 matrix $(e_i{}^\alpha)$ is nonsingular, its inverse $(e_\alpha{}^i)$ specifies a *tetrad field*, which belongs to the representation S of H:

$$\delta e_\alpha{}^i = \epsilon^b c_{\alpha b}{}^\gamma e_\gamma{}^i. \qquad (2.20)$$

We shall employ these matrices to convert holonomic indices $i,j,...$ to anholonomic indices $\alpha,\beta,...$ (which are associated with the representations S and $\widetilde{S}$ of H), in the usual way. For example, $\chi_i = e_i{}^\alpha\chi_\alpha$ and $\chi^i = \chi^\alpha e_\alpha{}^i$ are space-time vectors, invariant under H.

The third term in (2.19) shows that in general the $\Gamma_i{}^a$ are *not* Yang–Mills potentials for H; they are Yang–Mills potentials only if the condition (2.15) holds. Nevertheless, we shall employ the $\Gamma_i{}^a$ to define a *parallel transport* of a field, and an associated *generalized derivative*

$$D_i\psi = \partial_i\psi - \overline{\Gamma}_i\psi,\quad \overline{\Gamma}_i = \Gamma_i{}^a\overline{G}_a, \qquad (2.21)$$

which is not a true covariant derivative (transforming like ψ under the action of H) unless (2.15) holds. Nevertheless, it plays a crucial role in our gauge theory of the group G. We have shown elsewhere[15] how such a noncovariant derivative arises when the conformal group is gauged following Kibble's method of gauging the Poincaré group. In Sec. IV, we shall see that D_i is actually a constituent of a "generalized covariant derivative."

The transformation law of $D_i\psi$ under the action of H is easily found. It is

$$\delta D_i\psi = \epsilon^b(\overline{G}_b D_i\psi + e_i{}^\gamma c_{\gamma b}{}^a\overline{G}_a\psi). \qquad (2.22)$$

The rule for constructing the generalized derivative D_iX of a field variable is to subtract from ∂_iX the expression obtained by replacing ϵ^b by $-\Gamma_i{}^b$ in the infinitesimal change δX brought about by the action of H. Applying this rule, we find that

$$[D_i,D_j]\psi = -F_{ij}{}^a\overline{G}_a\psi, \qquad (2.23)$$

where

$$F_{ij}{}^a = \partial_i\Gamma_j{}^a - \partial_j\Gamma_i{}^a - \Gamma_i{}^b\Gamma_j{}^c c_{bc}{}^a$$
$$+ (e_j{}^\gamma\Gamma_i{}^b - e_i{}^\gamma\Gamma_j{}^b)c_{\gamma b}{}^a. \qquad (2.24)$$

This important quantity will be called *H-curvature*. Equally important is the *H-torsion* defined by

$$D_ie_j{}^\alpha - D_je_i{}^\alpha = F_{ij}{}^\alpha. \qquad (2.25)$$

That is,

$$F_{ij}{}^\alpha = \partial_ie_j{}^\alpha - \partial_je_i{}^\alpha + (e_j{}^\gamma\Gamma_i{}^b - e_i{}^\gamma\Gamma_j{}^b)c_{\gamma b}{}^\alpha. \qquad (2.26)$$

Comparison of (2.10) with (2.24) and (2.26) shows that the H-curvature and H-torsion are related to the field strengths for G through the relations

$$G_{ij}{}^A = F_{ij}{}^A - e_i{}^\beta e_j{}^\gamma c_{\beta\gamma}{}^A. \qquad (2.27)$$

In particular, if the π_α generate an Abelian subgroup of G (as is the case, for example, for the Poincaré group and conformal group), then $G_{ij}{}^A = F_{ij}{}^A$.

III. SPACE-TIME DIFFEOMORPHISMS

The Yang–Mills potentials $\Gamma_i{}^A$ transform as a covariant vector under general coordinate transformations or space-time diffeomorphisms. That is, under an infinitesimal space-time diffeomorphism $x^i\to x^i - \xi^i$, combined with an infinitesimal action of the group G,

$$\delta\Gamma_i{}^A = \xi^i\partial_j\Gamma_i{}^A + \Gamma_j{}^A\partial_i\xi^j + \nabla_i\epsilon^A \qquad (3.1)$$

[where δ denotes the substantial variation, $\delta X = X'(x) - X(x)$]. In terms of the new parameters

$$\lambda^A = \epsilon^A + \xi^i\Gamma_i{}^A, \qquad (3.2)$$

this is just

$$\delta\Gamma_i{}^A = \xi^jG_{ij}{}^A + \nabla_i\lambda^A. \qquad (3.3)$$

We now link the action of the infinitesimal generators π_α to the space-time diffeomorphisms by making the identification

$$\lambda^\alpha = \xi^\alpha. \qquad (3.4)$$

This step is of central importance in our approach. The gauge group G now has an action on the space-time points as well as on field components. Equation (3.3) now becomes

$$\delta\Gamma_i{}^A = \lambda^\beta F_{\beta i}{}^A + D_i\lambda^A + \lambda^b e_i{}^\gamma c_{b\gamma}{}^A, \qquad (3.5)$$

or, more explicitly,

$$\delta e_i{}^\alpha = \lambda^\beta F_{\beta i}{}^\alpha + \partial_i\lambda^\alpha + (\lambda^b e_i{}^\gamma - \lambda^\gamma\Gamma_i{}^b)c_{b\gamma}{}^\alpha, \qquad (3.6)$$
$$\delta\Gamma_i{}^a = \lambda^\beta F_{\beta i}{}^a + \partial_i\lambda^a + \lambda^b\Gamma_i{}^c c_{bc}{}^a$$
$$+ (\lambda^b e_i{}^\gamma - \lambda^\gamma\Gamma_i{}^b)c_{b\gamma}{}^a. \qquad (3.7)$$

This change is exactly the change brought about by a space-time diffeomorphism combined with an action of H [as is obvious from the fact that (3.4) is $\epsilon^\alpha = 0$]. It can therefore be associated with the change

$$\delta\psi = \xi^i\partial_i\psi + \epsilon^a\overline{G}_a\psi \qquad (3.8)$$

in a field ψ belonging to a linear representation of H, scalar under the diffeomorphisms. In terms of the new parameters (3.2), this is

$$\delta\psi = \lambda^\alpha D_\alpha\psi + \lambda^a\overline{G}_a\psi. \qquad (3.9)$$

In this form, we see that the action of the generators π_α is associated with *parallel transport* of ψ.

The transformation laws (3.5) *and* (3.9) *are the fundamental equations in our gauge theory of the group G.*

IV. THE MODIFIED LIE ALGEBRA

From (2.20) and (2.22) we can deduce the transformation law, under H, of the anholonomic generalized derivative $D_\alpha\psi = e_\alpha{}^iD_i\psi$; we find

$$\delta D_\alpha\psi = \epsilon^b(\overline{G}_bD_\alpha\psi + c_{\alpha b}{}^\gamma D_\gamma\psi + c_{\alpha b}{}^c\overline{G}_c\psi). \qquad (4.1)$$

Observe also that, under H,

$$\delta\overline{G}_a\psi=\overline{G}_a\delta\psi=\epsilon^b\overline{G}_a\overline{G}_b\psi=\epsilon^b(\overline{G}_b\overline{G}_a+c_{ab}{}^c G_c)\psi. \quad (4.2)$$

These two transformation laws can be combined in the single expression

$$\delta Q_A\psi=\epsilon^b(\overline{G}_b Q_A\psi+c_{Ab}{}^C Q_C\psi), \quad (4.3)$$

where the generators Q_A are defined by

$$Q_\alpha\psi=-D_\alpha\psi,\quad Q_a\psi=-\overline{G}_a\psi. \quad (4.4)$$

Thus, the components of $Q_A\psi$ transform according to the representation $T\otimes R$ of H.

Now, the Q_A are the operators that generate the changes (3.9) in ψ,

$$\delta\psi=-\lambda^A Q_A\psi. \quad (4.5)$$

We shall now look for the commutation relations satisfied by these operators.

The transformation law of $Q_A\psi$ under H is

$$\epsilon^b Q_b Q_A\psi=-\delta Q_A\psi=-\epsilon^b(\overline{G}_b Q_A\psi+c_{Ab}{}^c Q_c\psi).$$

Hence,

$$Q_b Q_a\psi=(\overline{G}_b\overline{G}_a+c_{ab}{}^c\overline{G}_c)\psi=\overline{G}_a\overline{G}_b\psi.$$

Therefore,

$$[Q_b,Q_a]\psi=[\overline{G}_a,\overline{G}_b]\psi=c_{ab}{}^c\overline{G}_c\psi=c_{ba}{}^c Q_c\psi,$$

establishing that

$$[Q_a,Q_b]=c_{ab}{}^c Q_c. \quad (4.6)$$

Also, from (4.3), we have

$$Q_b Q_\alpha\psi=G_b D_\alpha\psi-c_{ab}{}^C Q_C\psi. \quad (4.7)$$

Since the $\overline{G}_b$ are constant matrices, their generalized derivative vanishes

$$[D_i\overline{G}_b=\partial_i\overline{G}_b-\Gamma_i{}^a(\overline{G}_a\overline{G}_b-\overline{G}_b\overline{G}_a-c_{ab}{}^c\overline{G}_c)=0].$$

Therefore

$$Q_\alpha Q_b\psi=D_\alpha\overline{G}_b\psi=\overline{G}_b D_\alpha\psi. \quad (4.8)$$

Subtracting (5.8) from (5.7), we find that

$$[Q_\alpha,Q_b]=C_{ab}{}^\gamma Q_\gamma+c_{ab}{}^c Q_c. \quad (4.9)$$

The relations (4.6) and (4.9) for the Q_A are just like the commutation relations (2.1) with which we set out. However, the first commutator (2.1) is modified in a manner already familiar from Poincaré gauge theory.[4] We have

$$\begin{aligned}[D_\alpha,D_\beta]\psi&=e_\alpha{}^i D_i(\epsilon_\beta{}^j D_j\psi)-(\alpha\leftrightarrow\beta)\\&=(D_\alpha e_\beta{}^j-D_\beta e_\alpha{}^j)D_j\psi+e_\alpha{}^i e_\beta{}^j[D_i,D_j]\psi\\&=(D_\alpha e_\beta{}^j-D_\beta e_\alpha{}^j)D_j\psi-F_{\alpha\beta}{}^c\overline{G}_c\psi.\end{aligned}$$

But

$$\begin{aligned}F_{\alpha\beta}{}^j&=e_\gamma{}^j e_\alpha{}^i e_\beta{}^k(D_i e_k{}^\gamma-D_k e_i{}^\gamma)\\&=e_\gamma{}^j(e_\beta{}^k D_\alpha e_k{}^\gamma-e_\alpha{}^i D_\beta e_i{}^\gamma)=D_\beta e_\alpha{}^j-D_\alpha e_\beta{}^j.\end{aligned}$$

Therefore,

$$[D_\alpha,D_\beta]=-F_{\alpha\beta}{}^\gamma D_\gamma-F_{\alpha\beta}{}^c G_c, \quad (4.10)$$

which establishes that

$$[Q_\alpha,Q_\beta]=F_{\alpha\beta}{}^\gamma Q_\gamma+F_{\alpha\beta}{}^c Q_c. \quad (4.11)$$

The H-curvature and H-torsion have taken the place of structure constants, in a modification of the Lie algebra of G.

We shall now write the commutation relations for the Q_A, that we have just found, in the more concise form

$$[Q_A,Q_B]=F_{AB}{}^C Q_C, \quad (4.12)$$

where the $F_{\alpha\beta}{}^C$ are the H-curvature and H-torsion and the remaining components of $F_{AB}{}^C$ are the original structure constants of G. Observe that, if the curvature G_{ij} vanishes, the commutation relations (4.12) reduce to those of the Lie algebra of G [see (2.27)].

The fact that the appropriate derivative operator D_α for the gauge theory of G is not in general a covariant derivative operator is at first sight disturbing. However, recall that the covariant derivative operator associated with a gauge group H is by definition an operator that acts on a field ψ, belonging to a linear homogeneous representation R of H, to produce a derivative of H that transforms linearly and homogeneously. In the present context, such a covariant derivative operator does in fact exist, namely the operator

$$-Q_A=\begin{pmatrix}D_\alpha\\\overline{G}_a\end{pmatrix}, \quad (4.13)$$

which produces a derivative transforming according to the linear homogeneous representation $T\otimes R$. Thus, we have an interesting extension of the usual notion of covariant differentiation.

Now let ϕ_A and ϕ^A be quantities belonging to the representations T and $\widetilde{T}$ of H. We define a new derivative operator for such quantities, suggested by the relations (4.12) and the usual structure of "covariant derivatives of the adjoint and coadjoint representations of a Lie group":

$$\begin{aligned}\mathscr{D}_i\phi_A&=\partial_i\phi_A-\Gamma_i{}^B F_{AB}{}^C\phi_C,\\\mathscr{D}_i\phi^A&=\partial_i\phi^A+\phi^C\Gamma_i{}^B F_{CB}{}^A.\end{aligned} \quad (4.14)$$

Unlike the quantities $D_i\phi_A$ and $D_i\phi^A$, we find that these derivatives are true covariant derivatives in that they transform like ϕ_A (resp. ϕ^A) under the action of H. The geometrical significance of the operator $\mathscr{D}_i$ is at present obscure. However, as we shall see, it leads to striking formal simplifications of some of the fundamental relationships of our theory.

V. THE SOURCE IDENTITIES

Any gauge theory has two distinct aspects: the purely geometrical aspect and the physical aspect. The physics is introduced by means of Lagrangians, and, in the case of gauge theories of space-time groups, by means of hypotheses concerning the relationship between the potentials and the metric and affine properties of space-time. In the preceding sections, we have set up the formalism for the geometrical aspect of a gauge theory of the group G. We now relate this to physics by postulating the existence of Lagrangian theories invariant under the action of the gauge transformations.

Suppose there exists a Lagrangian density $\mathscr{L}(\psi,\partial_i\psi,\Gamma_i{}^A)$ whose field equations are form-invariant under the action of the transformations (3.5) and (3.9). The interesting question of what are the possible forms of Lagrangian densities (if any) for a given group G will not be dealt with here; we simply assume the existence of $\mathscr{L}$ and examine the consequences of that assumption.

The covariance requirement is

$$\partial_i(\xi^i \mathscr{L}) = \delta \mathscr{L} = \frac{\partial \mathscr{L}}{\partial \psi}\delta\psi + \frac{\partial \mathscr{L}}{\partial\, \partial_i \psi}\delta \partial_i \psi + \frac{\partial \mathscr{L}}{\partial \Gamma_i{}^A}\delta \Gamma_i{}^A. \tag{5.1}$$

Define the *sources* of the Yang–Mills potentials

$$\Sigma_A{}^i = \frac{\partial \mathscr{L}}{\partial \Gamma_i{}^A} \tag{5.2}$$

and employ the field equations

$$0 = \frac{\delta \mathscr{L}}{\delta \psi} = \frac{\partial \mathscr{L}}{\partial \psi} - \partial_i \Pi^i, \quad \Pi^i = \frac{\partial \mathscr{L}}{\partial\, \partial_i \psi}. \tag{5.3}$$

We find that

$$\partial_i(\xi^i \mathscr{L} - \Pi^i \delta\psi) = \Sigma_A{}^i \delta \Gamma_i{}^A. \tag{5.4}$$

Now,

$$\xi^i \mathscr{L} - \Pi^i \delta\psi = \lambda^A \theta_A{}^i, \tag{5.5}$$

where

$$\theta_\alpha{}^i = \mathscr{L} e_\alpha{}^i - \Pi^i D_\alpha \psi, \tag{5.6}$$

$$\theta_a{}^i = -\Pi^i \overline{G}_a \psi. \tag{5.7}$$

These quantities are recognizable as a canonical energy-momentum density and a set of (intrinsic) currents associated with the subgroup H. We find that $\theta_A{}^i$ belongs to the representation T of H,

$$\delta \theta_A{}^i = \lambda^b c_{Ab}{}^C \theta_C{}^i \tag{5.8}$$

(and is a vector density under space-time diffeomorphisms).

We now have

$$\partial_i(\lambda^A \theta_A{}^i) = D_i(\lambda^A \theta_A{}^i) = \Sigma_A{}^i(\lambda^\alpha F_{\alpha i}{}^A + D_i \lambda^A + \lambda^b e_i{}^\gamma c_{b\gamma}{}^A). \tag{5.9}$$

Equating coefficients of $D_i \lambda^A$ identifies the sources of the Yang–Mills potentials as the canonical currents,

$$\Sigma_A{}^i = \theta_A{}^i. \tag{5.10}$$

Equating coefficients of λ^α and λ^a gives the source identities

$$D_i \theta_\alpha{}^i = \theta_A{}^i F_{\alpha i}{}^A, \tag{5.11}$$

$$D_i \theta_a{}^i = \theta_A{}^i c_{ai}{}^A. \tag{5.12}$$

The first of these is recognizable as a generalized energy-momentum conservation law, the right-hand side being a set of "Lorentz forces" constructed from the currents and field strengths.

The relations (5.10) are explicitly

$$\frac{\partial \mathscr{L}}{\partial e_i{}^\alpha} = \mathscr{L} e_\alpha{}^i - \frac{\partial \mathscr{L}}{\partial\, \partial_i \psi} D_\alpha \psi \tag{5.13}$$

and

$$\frac{\partial \mathscr{L}}{\partial \Gamma_i{}^a} = -\frac{\partial \mathscr{L}}{\partial\, \partial_i \psi} \overline{G}_a \psi. \tag{5.14}$$

They imply that the Lagrangian density has the form

$$\mathscr{L} = eL(\psi, D_\alpha \psi), \quad e = |e_i{}^\alpha|. \tag{5.15}$$

VI. THE FIELD EQUATIONS

Let us now suppose the existence of a Lagrangian density $\mathscr{V}(\Gamma_i{}^A, \partial_j \Gamma_i{}^A)$ for the Yang–Mills potentials, and add it to $\mathscr{L}$. Covariance requirements impose the restriction

$$\delta \mathscr{V} = \partial_i(\xi^i \mathscr{V}) = \frac{\partial \mathscr{V}}{\partial \Gamma_i{}^A}\delta \Gamma_i{}^A + \mathscr{H}_A{}^{ij} \delta\, \partial_j \Gamma_i{}^A, \tag{6.1}$$

where

$$\mathscr{H}_A{}^{ij} = \frac{\partial \mathscr{V}}{\partial_j \Gamma_i{}^A}. \tag{6.2}$$

(The possibility of constructing such Lagrangians will be considered in Sec. VII.)

The field equations obtained from variation of $\Gamma_i{}^A$ in $\mathscr{L} + \mathscr{V}$ are

$$\frac{\delta \mathscr{V}}{\delta \Gamma_i{}^A} = \frac{\partial \mathscr{V}}{\partial \Gamma_i{}^A} - \partial_j \mathscr{H}_A{}^{ij} = -\theta_A{}^i. \tag{6.3}$$

Hence

$$\partial_j(\xi^j \mathscr{V} - \mathscr{H}_A{}^{ij} \delta \Gamma_i{}^A) = -\theta_A{}^i \delta \Gamma_i{}^A, \tag{6.4}$$

i.e.,

$$D_j[\lambda^A \mathscr{E}_A{}^j - \mathscr{H}_A{}^{ij} D_i \lambda^A] = -\theta_A{}^i(\lambda^\beta F_{\beta i}{}^A + D_i \lambda^A + \lambda^b e_i{}^\gamma c_{b\gamma}{}^A), \tag{6.5}$$

where

$$\mathscr{E}_\alpha{}^j = \mathscr{V} e_\alpha{}^j - \mathscr{H}_A{}^{\beta j} F_{\alpha\beta}{}^A, \tag{6.6}$$

$$\mathscr{E}_a{}^j = -\mathscr{H}_A{}^{\beta j} c_{a\beta}{}^A. \tag{6.7}$$

Equating coefficients of $\partial_i\, \partial_j \lambda^A$ gives

$$\mathscr{H}_A{}^{ij} = -\mathscr{H}_A{}^{ji} \tag{6.8}$$

(which shows that derivatives of the $\Gamma_i{}^A$ have to be contained in $\mathscr{V}$ in the combination $G_{ij}{}^A$), and equating coefficients of $D_j \lambda^A$ gives the field equations for the Yang–Mills potentials in the Maxwellian form

$$D_j \mathscr{H}_A{}^{ij} = \theta_A{}^i + \mathscr{E}_A{}^i. \tag{6.9}$$

Observe that the $\mathscr{E}_A{}^i$ are the energy-momentum density and H-currents for the Yang–Mills potentials. Equation (5.4) can be regarded as a definition of the energy-momentum and the H-currents. Applying this to the Yang–Mills Lagrangian, we find

$$\xi^j \mathscr{V} - \frac{\partial \mathscr{V}}{\partial\, \partial_j \Gamma_i{}^A}\delta \Gamma_i{}^A = \lambda^A \mathscr{E}_A{}^j + \text{terms in } D_i \lambda^A. \tag{6.10}$$

The peculiar derivative operator $\mathscr{D}_i$ introduced in Sec. IV can be used to cast some of our equations into a particularly elegant form. For example, the transformation law (3.5) for the Yang–Mills potentials is just

$$\delta \Gamma_i{}^A = \mathscr{D}_i \lambda^A, \tag{6.11}$$

the source identities (5.11) and (5.12) are

$$\mathscr{D}_i \theta_A{}^i = 0 \tag{6.12}$$

and the field equations (6.9) are

$$\mathscr{D}_i \mathscr{H}_A{}^{ij} = \theta_A{}^i + e_A{}^i \mathscr{V} \tag{6.13}$$

(where $e_a{}^i$ is defined to be zero). Note that the expressions on the right and left in (6.12) and (6.13) transform linearly and *homogeneously* under the action of H (they belong to the representation T), but this was not the case for the source identities and field equations as originally given, unless $c_{ab}{}^{c}=0$. Note finally that the definitions (5.6), (5.7), (6.6), and (6.7) of energy-momentum densities and H-currents can be written in the "manifestly covariant" forms

$$\theta_A{}^i = \mathscr{L} e_A{}^i + \Pi^i Q_A \psi, \tag{6.14}$$

$$\mathscr{E}_A{}^i = \mathscr{V} e_A{}^i - \mathscr{H}_B{}^{\beta i} F_{A\beta}{}^B. \tag{6.15}$$

VII. STRUCTURE OF LAGRANGIANS

So far, we have not proposed any particular form for the Lagrange density $\mathscr{V}$. We have found that it has to be a function of $\Gamma_i{}^A$ and $G_{ij}{}^A$. An obvious choice is the Maxwell-type Lagrange density quadratic in curvature,

$$\mathscr{V} = (1/\kappa) g^{1/2} g^{ij} g^{kl} G_{ik}{}^A G_{jl}{}^B \gamma_{AB}, \tag{7.1}$$

where κ is a constant, the matrix (g_{ij}) is constructed from the tetrad according to

$$g_{ij} = e_i{}^\alpha e_j{}^\beta \eta_{\alpha\beta}, \tag{7.2}$$

where $(\eta_{\alpha\beta})$ is some nonsingular symmetric matrix, (g^{ij}) is the inverse of (g_{ij}), and g is its determinant. The matrix (γ_{AB}) is the Cartan form for G,

$$\gamma_{AB} = -c_{EA}{}^F c_{FB}{}^E. \tag{7.3}$$

Clearly, g_{ij} transforms as a tensor under coordinate transformations or diffeomorphisms of space-time (and can be interpreted as the space-time metric). The function (7.1) is a scalar density under coordinate transformations or space-time diffeomorphisms. It will be invariant under the action of H provided

$$c_{\alpha b}{}^{\gamma}\eta_{\gamma\beta} + c_{\beta b}{}^{\gamma}\eta_{\gamma\alpha} = \tfrac{1}{2}\eta_{\alpha\beta} c_{\gamma b}{}^{\gamma}. \tag{7.4}$$

This can be regarded as a restriction on the choice of the group G when η is given. In terms of the four-dimensional representation S of H, it can be written more succinctly as

$$S\eta S^T = |S|^{1/2}\eta. \tag{7.5}$$

In the following section we shall show that the gauge theory of the group G that we have set up can be obtained by gauging a global group of space-time diffeomorphisms. This will of course only lead to plausible physics if the global group of diffeomorphisms is related to the geometrical properties of the space-time on which it acts (e.g., Poincaré or conformal transformations on Minkowski space, de Sitter transformations on de Sitter space). The Poincaré gauge theory, de Sitter gauge theory, and conformal gauge theory all have groups G that satisfy a condition of the form (7.5). The affine group does not, but in the affine gauge theory[8] an independent dynamical metric field g_{ij}, unrelated to the tetrad, is introduced into the structure of Lagrangians. Of course, in some cases (7.1) will not be the uniquely possible choice for $\mathscr{V}$. For example, as is well known, in Poincaré gauge theory a curvature scalar is a possible choice,[5] and so are terms quadratic in torsion.[4]

The possible forms for the matter Lagrangian present a more complicated problem, which will not be dealt with here, except to mention that matter Lagrangians with the required transformation laws can be constructed for the gauge theories of the Poincaré, de Sitter, and conformal groups.

The gauge theories of the de Sitter and conformal groups that arise as particular cases of our formalism will be presented in a sequel to the present work.

VIII. REALIZATION OF G AS A GLOBAL GROUP OF DIFFEOMORPHISMS

We consider now what happens when the curvature G_{ij} vanishes. We have noted already that the operators Q_A then satisfy the commutation relations of the Lie algebra of G. Let

$$\mathring{\Gamma}_i = \Gamma_i{}^A G_A \tag{8.1}$$

be a connection, with vanishing curvature, whose coefficients are given functions of space-time (note that $\mathring{e}_i{}^\alpha$ is required to be a nonsingular matrix, so $\mathring{\Gamma}_i = 0$ is not an appropriate choice). The transformations (3.5) that preserve the given form of the $\mathring{\Gamma}_i{}^A$ are

$$\nabla_i \lambda^A = 0 = \partial_i \lambda^A + \lambda^C \mathring{\Gamma}_i{}^D c_{CD}{}^A. \tag{8.2}$$

We have a set of $N+4$ linear differential equations to be solved for the $N+4$ parameters λ^A. The integrability conditions are $\mathring{G}_{ij}{}^A = 0$, which are satisfied. The general solution can be written in terms of an element σ of the group G, satisfying

$$\partial_i \sigma + \sigma \mathring{\Gamma}_i = 0. \tag{8.3}$$

The integrability conditions for these equations are $G_{ij} = 0$. Now denote the matrix that represents σ in the adjoint representation by $E_B{}^A$. Then

$$\partial_i E_B{}^A + E_B{}^C \mathring{\Gamma}_i{}^D c_{CD}{}^A = 0. \tag{8.4}$$

Since the matrix $E_B{}^A$ is nonsingular, its columns provide $N+4$ linearly independent solutions of (8.2). The general solution is therefore

$$\lambda^A = a^B E_B{}^A, \tag{8.5}$$

where the a^B are constants.

The transition law of ψ, under these specialized transformations, is

$$\delta\psi = a^B M_B, \tag{8.6}$$

$$M_B = E_B{}^\alpha D_\alpha + E_B{}^a \overline{G}_a. \tag{8.7}$$

In order to establish that (8.6) corresponds to a representation of G, with a^B as parameters, we have to show that

$$[M_A, M_B] = c_{AB}{}^C M_C. \tag{8.8}$$

Let $\lambda^A = a^B E_B{}^A$ and $\mu^A = b^B E_B{}^A$ be two arbitrary solutions of (8.2). Then, since a^B and b^B are constants, we have

$$\begin{aligned} a^A b^B [M_A, M_B] &= [\lambda^\alpha D_\alpha + \lambda^a \overline{G}_a, \mu^\beta D_\beta + \mu^b \overline{G}_b] \\ &= (\lambda^\alpha D_\alpha \mu^\beta - (\lambda \leftrightarrow \mu)) D_\beta + \lambda^\alpha \mu^\beta [D_\alpha, D_\beta] \\ &\quad + (\lambda^\alpha D_\alpha \mu^b - (\lambda \leftrightarrow \mu)) \overline{G}_b + \lambda^a \lambda^b [\overline{G}_a, \overline{G}_b]. \end{aligned}$$

Now, (8.2) can be written

$$D_\alpha \lambda^A = -\lambda^B c_{B\alpha}{}^A, \tag{8.9}$$

and we have an identical equation in μ^A. Therefore

$$a^A b^B [M_A, M_B] = \lambda^A \mu^B (c_{AB}{}^\gamma D_\gamma + c_{AB}{}^c \overline{G}_c)$$

(in arriving at this, we make use of $c_{ab}{}^{\gamma}=0$). Therefore

$$[M_A,M_B]=E_A{}^E E_B{}^F(c_{EF}{}^{\gamma}D_\gamma+c_{EF}{}^c\overline{G}_c).$$

Since $E_A{}^B$ belongs to the adjoint representation of G, it satisfies the identity

$$E_A{}^E E_B{}^F c_{EF}{}^C=c_{AB}{}^D E_D{}^C. \tag{8.10}$$

The result (8.8) then follows immediately.

An alternative form for the M_B is

$$M_B=B_B{}^i\partial_i+B_B{}^a\overline{G}_a, \tag{8.11}$$

where the coefficients are defined by

$$E_B{}^\alpha=B_B{}^i\mathring{e}_i{}^\alpha,\quad E_B{}^a=B_B{}^a+B_B{}^i\mathring{\Gamma}_i{}^a. \tag{8.12}$$

The commutation relations (8.8) imply the following identities:

$$B_A{}^i\partial_i B_B{}^j-B_B{}^i\partial_i B_A{}^j=c_{AB}{}^C B_C{}^j, \tag{8.13}$$

$$B_A{}^i\partial_i B_B{}^a-B_B{}^i\partial_i B_A{}^a=c_{AB}{}^C B_C{}^a-B_A{}^e B_B{}^f c_{ef}{}^a. \tag{8.14}$$

The first of these relations shows that the group G is now realized as a global group of diffeomorphisms $x^i\to x^i-\xi^i$, with

$$\xi^i=a^B B_B{}^i. \tag{8.15}$$

The relations (8.14) were given by Harnad and Pettitt[7] as the necessary conditions for the transformation law (8.6) [with M_B given by (8.11)] to represent a global group of diffeomorphisms of the form (8.15).

The procedure adopted in this section is the inverse of the usual one—we started with a gauge theory and "ungauged it," ending up with a global group of transformations. The Lagrangian density (5.15), if it exists, gives rise to a Lagrangian density

$$\mathscr{L}(\psi,\partial_i\psi,x^i)=\mathring{e}L(\psi,\mathring{D}_\alpha\psi), \tag{8.16}$$

where

$$\mathring{D}_\alpha\psi=\mathring{e}_\alpha{}^i(\partial_i\psi-\mathring{\Gamma}_i{}^a\overline{G}_a\psi), \tag{8.17}$$

for a theory that is covariant under a global group of diffeomorphisms (the $\mathring{\Gamma}_i{}^A$ are invariant specified functions of the coordinates, not fields). The Noether currents for this theory are defined by

$$\xi^i\mathscr{L}-\Pi^i\delta\psi=a^B J_B{}^i. \tag{8.18}$$

They are

$$J_B{}^j=B_B{}^i\hat{\theta}_i{}^j+B_B{}^A\theta_a{}^j, \tag{8.19}$$

where

$$\hat{\theta}_i{}^j=\mathscr{L}\delta_i^j-\Pi^j\partial_i\psi, \tag{8.20}$$

$$\theta_a{}^j=-\Pi^j\overline{G}_a\psi. \tag{8.21}$$

They satisfy

$$\partial_j J_B{}^j=0. \tag{8.22}$$

Alternatively, the Noether currents can be written in the form

$$J_B{}^j=E_B{}^A\theta_A{}^j, \tag{8.23}$$

where

$$\theta_\alpha{}^j=\mathscr{L}\mathring{e}_\alpha{}^j-\Pi^j\mathring{D}_\alpha\psi. \tag{8.24}$$

The conservation laws (8.22) then take the form

$$\partial_j\theta_A{}^j=\mathring{\Gamma}_j{}^B c_{AB}{}^C\theta_C{}^j, \tag{8.25}$$

which are of course the limiting cases of the source identities (5.11). The quantities $\theta_A{}^j$ are "intrinsic" currents for the group G and the $J_A{}^j$ are the "total" (intrinsic + orbital) currents.

The covariance of (8.16) is of course lost when the parameters a^A are made space-time dependent [which is tantamount to making the λ^A independent space-time-dependent functions by abandoning the constraint (8.2)]. The change in the Lagrangian is now

$$\delta\mathscr{L}=\partial_i(\xi^i\mathscr{L})-(\partial_i a^B)J_B{}^i. \tag{8.26}$$

Obviously, the covariance can be maintained by introducing auxiliary fields $\Gamma_i{}^A$ so as to revert to the original theory of Secs. V and VI.

Important particular cases of the foregoing theory arise when the π_α commute ($c_{\alpha\beta}{}^C=0$). The Poincaré group gauge theories and conformal gauge theory belong to this class. An appropriate choice for the $\mathring{\Gamma}_i{}^A$ in these cases is

$$\mathring{e}_i{}^\alpha=\delta_i^\alpha,\quad \mathring{\Gamma}_i{}^a=0. \tag{8.27}$$

The distinction between Latin and Greek indices, and the distinction between the generalized derivative and the ordinary partial derivative, now disappear. The constraint (8.2) on the transformation parameters is now

$$\partial_\gamma\lambda^A+\lambda^B c_{B\gamma}{}^A=0. \tag{8.28}$$

The matrix σ that solves (7.3) is

$$\sigma=e^{-\pi\cdot x},\quad \pi\cdot x=\pi_\alpha x^\alpha, \tag{8.29}$$

so that

$$E_B{}^A=(e^{-c\cdot x})_B{}^A,\quad c\cdot x=c_\alpha x^\alpha, \tag{8.30}$$

where the four matrices c_α are the adjoint representatives of the π_α,

$$(c_\alpha)_B{}^A=c_{B\alpha}{}^A. \tag{8.31}$$

On account of $c_{\alpha\gamma}{}^B=0$, we have

$$E_\alpha{}^\beta=\delta_\alpha^\beta,\quad E_\alpha{}^b=0, \tag{8.32}$$

and consequently

$$\xi^\alpha=a^\alpha+a^b E_b{}^\alpha \tag{8.33}$$

and

$$\delta\psi=a^\alpha\partial_\alpha\psi+a^b(E_b{}^\alpha\partial_\alpha+E_b{}^a\overline{G}_a)\psi. \tag{8.34}$$

The Noether currents in these cases are

$$J_\beta{}^j=\theta_\beta{}^j,\quad J_b{}^j=E_b{}^\alpha\theta_\alpha{}^j+E_b{}^a\theta_a{}^j. \tag{8.35}$$

The two pieces of the right-hand side of the final expression correspond to the "intrinsic" and "orbital" parts of the current.

ACKNOWLEDGMENT

We wish to thank the University Grants Commission of India for the financial support of this work.

[1] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).
[2] R. Utiyama, Phys. Rev. **101**, 1597 (1956).
[3] D. W. Sciama, in *Recent Developments in General Relativity*, Festschrift for Infeld (Pergamon, New York, 1962).
[4] F. W. Hehl, *Four Lectures on Poincaré Gauge Field Theory*, in *Proceedings*

of the 6th Course of the International School of Cosmology and Gravitation, edited by P. G. Bergmann and V. de Sabbata (Plenum, New York, 1978).
[5]T. W. B. Kibble, J. Math. Phys. **2**, 212 (1961).
[6]P. von der Heyde, Phys. Lett. A **58**, 141 (1976).
[7]J. P. Harnad and R. B. Pettitt, J. Math. Phys. **17**, 1827 (1976).
[8]E. A. Lord, Phys. Lett. A **65**, 1 (1978).
[9]F. W. Hehl, G. D. Kerlick, and P. von der Heyde, Phys. Lett. B **63**, 446 (1976); F. W. Hehl, G. D. Kerlick, E. A. Lord, and L. L. Smalley, Phys. Lett. B **70**, 70 (1977); F. W. Hehl and G. W. Kerlick, Gen. Relativ. Gravit. **9**, 691 (1978); F. W. Hehl, E. A. Lord, and L. L. Smalley, *ibid.* **13**, 1037 (1981).
[10]F. W. Hehl, E. A. Lord, and Y. Ne'eman, Phys. Rev. D **17**, 428 (1978); Phys. Lett. B **71**, 432 (1977); Y. Ne'eman and D. Sijacki, Ann. Phys. (NY) **120**, 292 (1979).
[11]S. MacDowell and R. Mansouri, Phys. Rev. Lett. **38**, 739 (1977).
[12]J. P. Harnad and R. B. Pettitt, in *Group Theoretical Methods in Physics, Proceedings of the V International Colloquium*, edited by R. T. Sharp and B. Kolman (Academic, New York, 1977).
[13]K. P. Sinha, Pramana **23**, 205 (1984).
[14]D. Ivanenko and Sardanashvily, Phys. Rep. **94**, 1 (1983).
[15]E. A. Lord and P. Goswami, Pramana **25**, 635 (1985).

Scalar formalism for non-Abelian gauge theory

Levere C. Hostler[a)]
Newman Laboratory of Nuclear Studies, Cornell University, Ithaca, New York 14853-5001

(Received 23 September 1985; accepted for publication 30 April 1986)

The gauge field theory of an N-dimensional multiplet of spin-$\frac{1}{2}$ particles is investigated using the Klein–Gordon-type wave equation $\{\Pi\cdot(1+i\sigma)\cdot\Pi+m^2\}\Phi=0$, $\Pi_\mu\equiv\partial/\partial ix_\mu - eA_\mu$, investigated before by a number of authors, to describe the fermions. Here Φ is a 2×1 Pauli spinor, and σ represents a Lorentz spin tensor whose components $\sigma_{\mu\nu}$ are ordinary 2×2 Pauli spin matrices. Feynman rules for the scalar formalism for non-Abelian gauge theory are derived starting from the conventional field theory of the multiplet and converting it to the new description. The equivalence of the new and the old formalism for arbitrary radiative processes is thereby established. The conversion to the scalar formalism is accomplished in a novel way by working in terms of the path integral representation of the generating functional of the vacuum τ-functions, $\tau(2,1,\cdots 3\cdots)\equiv\langle 0-|T(\Psi_{\text{in}}(2)\overline{\Psi}_{\text{in}}(1)\cdots A_\mu(3)_{\text{in}}\cdots S)|0-\rangle$, where Ψ_{in} is a Heisenberg operator belonging to a $4N\times1$ Dirac wave function of the multiplet. The Feynman rules obtained generalize earlier results for the Abelian case of quantum electrodynamics.

I. INTRODUCTION

The use of the "second-order" Dirac equation[1]

$$\{\Pi\cdot(1+i\sigma)\cdot\Pi+m^2\}\Phi=0\,,\quad \Pi_\mu\equiv\frac{\partial}{\partial ix_\mu}-eA_\mu\,, \tag{1.1}$$

to describe a spin-$\frac{1}{2}$ particle has been investigated earlier.[2–14] In Eq. (1.1) Φ is a 2×1 Pauli spinor, and σ is a spin tensor whose Lorentz components $\sigma_{\mu\nu}$ are ordinary 2×2 Pauli spin matrices [see Eq. (2.16)]. The use of Eq. (1.1) to describe a spin-$\frac{1}{2}$ particle brings out a close parallel between the quantum theory of a spin-$\frac{1}{2}$ particle and the quantum theory of a simple scalar particle. Indeed, Feynman rules for a quantum electrodynamics based on Eq. (1.1) are essentially the rules of scalar electrodynamics, aside from a modification of the one-photon vertex to incorporate spin effects. These Feynman rules have been derived using the c-number formalism[4] and again from the formalism of quantum field theory using a variety of approaches.[5,14] The similarity between this "scalar formalism" for quantum electrodynamics and scalar electrodynamics extends to the types of divergent graphs and to the renormalization prescriptions as well.

We will here extend the Feynman rules for quantum electrodynamics in the scalar formalism to include a non-Abelian gauge theory of an N-dimensional multiplet of fermions. Our method will be to start with the conventional gauge field theory of the multiplet employing the linear Dirac equation and to then convert to the new formalism, thereby establishing for arbitrary radiative processes the equivalence of the new and the old formalism.

In Sec. II a unitary transformation of the conventional linear Dirac equation for the fermion multiplet is performed, enabling us to write the Dirac equation of the multiplet in a second-order form, Eq. (2.14). This second-order form parallels closely Eq. (1.1). Accordingly, the derivation of the wave equation (2.17) of the dual state of the fermion multiplet and the derivation of a conserved transition current, Eq. (2.18), can be carried out just as in the simple one-particle case.

The new representation of the fermion multiplet derived in Sec. II is employed in Sec. III, where the Dirac field describing the multiplet is second quantized in the usual way with path integral techniques. Conversion to the language of the scalar formalism is accomplished by first formally performing the path integral over the fermion degrees of freedom. This step is made possible by using an abstract operator notation due to Schwinger.[15] The resulting proof of equivalence is quite striking for its simplicity and generality. The Feynman rules obtained generalize earlier results for the Abelian case of quantum electrodynamics.[4,5,14] In the Appendix the equivalence proof is reconsidered briefly using the method of Ref. 14.

The model considered is one without symmetry breaking. The Feynman rules, summarized in Sec. III, Table I, are basically the rules for a corresponding multiplet of scalar particles; except for a modification of the "gluon–quark" vertices to incorporate spin effects. Such vertex modifications are needed in the non-Abelian case for both the simple gluon–quark vertex and for the double gluon–quark vertex (seagull vertex); in contrast to the quantum electrodynamics case in which all spin effects reside in the one-photon vertex.

The motivation for this work stems from an interest in seeing a familiar theory recast in a new language. This new language may be of some heuristic and/or calculational value for elementary particle physics. A Bethe–Salpeter equation for two interacting fermions provides an example in which the scalar formalism may have an advantage over earlier methods, since the Bethe–Salpeter wave function for the two fermion system is a 2×2 matrix in the scalar formalism, as opposed to a 4×4 matrix in a description based on the familiar first-order Dirac equation.

II. SECOND-ORDER DIRAC EQUATION FOR AN N-DIMENSIONAL MULTIPLET OF FERMIONS

As indicated in the Introduction, our starting point will be the conventional gauge theory of an N-dimensional multiplet of fermions. The first-order Dirac equation for such a system reads

[a)] Permanent address: Physics Department, Wilkes College, Wilkes Barre, Pennsylvania 18766.

 0022-2488/86/092423-06$02.50

$$\{\Gamma \cdot (-i\partial - eA) - im\}\Psi = 0\,, \tag{2.1}$$

where

$$\Psi = \begin{bmatrix} \Psi_1 \\ \vdots \\ \Psi_N \end{bmatrix} \tag{2.2}$$

is an $N \times 1$ matrix of 4×1 one-particle Dirac spinors. The transition to the desired second-order form of Eq. (2.1) is brought about most simply by working in a representation for which the one-particle Dirac matrix γ_5 is diagonal. Such a representation of the one-particle Dirac matrices is obtained by choosing

$$\gamma \equiv \left[\begin{array}{c|c} 0 & i\sigma \\ \hline -i\sigma & 0 \end{array}\right], \quad \gamma_4 = \beta \equiv \left[\begin{array}{c|c} 0 & 1 \\ \hline 1 & 0 \end{array}\right]. \tag{2.3}$$

In Eq. (2.1) the Γ_μ are $N \times N$ matrices of 4×4 matrices. When written out in block form with 2×2 blocks, they have the following appearance:

$$\Gamma = \left[\begin{array}{cc|cc|c} 0 & i\sigma & & & \\ -i\sigma & 0 & & & \\ \hline & & 0 & i\sigma & \\ & & -i\sigma & 0 & \\ \hline & & & & \ddots \end{array}\right],$$

$$\Gamma_4 = \left[\begin{array}{cc|cc|c} 0 & 1 & & & \\ 1 & 0 & & & \\ \hline & & 0 & 1 & \\ & & 1 & 0 & \\ \hline & & & & \ddots \end{array}\right]. \tag{2.4}$$

Taking our cue from Ref. 14 we split each one-particle Dirac spinor in Eq. (2.2) into two 2×1 blocks as follows:

$$\Psi_a \equiv \begin{bmatrix} \Phi_a \\ \overline{\Phi}{}_a^\dagger/m \end{bmatrix}, \tag{2.5}$$

thereby defining the Pauli spinors Φ_a, $a = 1,2,\ldots,N$, and dual spinors $\overline{\Phi}_a$. Equation (2.2) now takes the form

$$\Psi = \begin{bmatrix} \Phi_1 \\ \overline{\Phi}{}_1^\dagger/m \\ \vdots \\ \Phi_N \\ \overline{\Phi}{}_N^\dagger/m \end{bmatrix}, \tag{2.6}$$

in which the spinors Φ alternate with their duals $\overline{\Phi}$. Next a canonical transformation is performed in which all spinors Φ in Eq. (2.6) are brought to the top. The transformed state Ψ^T will have the form

$$\Psi^T = \begin{bmatrix} \Phi_1 \\ \vdots \\ \Phi_N \\ \overline{\Phi}{}_1^\dagger/m \\ \vdots \\ \overline{\Phi}{}_N^\dagger/m \end{bmatrix} \equiv \begin{bmatrix} \Phi \\ \overline{\Phi}{}^\dagger/m \end{bmatrix}. \tag{2.7}$$

The expressions for the transforms $\Gamma_\mu{}^T$ of the matrices Γ_μ are

$$\Gamma^T = \left[\begin{array}{cccc|cccc} & & & & i\sigma & & & \\ & & & & & i\sigma & & \\ & & & & & & \ddots & \\ & & & & & & & i\sigma \\ \hline -i\sigma & & & & & & & \\ & -i\sigma & & & & & & \\ & & \ddots & & & & & \\ & & & -i\sigma & & & & \end{array}\right],$$

$$\Gamma_4^T = \left[\begin{array}{cccc|cccc} & & & & 1 & & & \\ & & & & & 1 & & \\ & & & & & & \ddots & \\ & & & & & & & 1 \\ \hline 1 & & & & & & & \\ & 1 & & & & & & \\ & & \ddots & & & & & \\ & & & 1 & & & & \end{array}\right]. \tag{2.8}$$

In terms of Kronecker products we have the relations

$$\Gamma_\mu = \mathbf{1}_{N\times N} \otimes \gamma_\mu$$

and

$$\Gamma_\mu{}^T = \gamma_\mu \otimes \mathbf{1}_{N\times N}.$$

We may write the matrices (2.8) in an abbreviated form

$$\Gamma^T = \left[\begin{array}{c|c} 0 & i\sigma \\ \hline -i\sigma & 0 \end{array}\right], \quad \Gamma_4^T = \left[\begin{array}{c|c} 0 & 1 \\ \hline 1 & 0 \end{array}\right], \tag{2.9}$$

in which factors of the $N \times N$ unit matrix are understood. All subsequent occurrences of Ψ, $\overline{\Psi}$, and Γ_μ shall refer to the transformed quantities. Note the similarity in structure between the expressions (2.9) for the $\Gamma_\mu{}^T$ and the expressions (2.3) for their one-particle counterparts γ_μ. This similarity in structure makes it possible to work with the fermion multiplet essentially as we did before in the one-particle case. Thus in terms of the variables Φ and $\overline{\Phi}$ of Eq. (2.7) our field equation (2.1) takes a familiar form

$$(-\vec{A}\vec{B} + m^2)\Phi = 0 \tag{2.10}$$

and

$$\overline{\Phi} = \Phi^\dagger \overleftarrow{B}\,, \tag{2.11}$$

in which

$$A = -i\Pi_4 + \sigma\cdot\Pi \equiv \bar{\tau}_\mu \Pi_\mu\,, \tag{2.12}$$

and

$$B = -i\Pi_4 - \sigma\cdot\Pi \equiv -\tau_\mu \Pi_\mu\,. \tag{2.13}$$

The system of equations (2.10) and (2.11) is entirely equivalent to the original field equation (2.1) from which we started and forms the foundation of our scalar formalism for the fermion multiplet. The parallel with the Klein–Gordon equation is brought out most clearly by writing out Eq. (2.10) more explicitly:

$$\{\Pi\cdot(1+i\sigma)\cdot\Pi+m^2\}\Phi=0\,, \tag{2.14}$$

a form that exploits the identity[16]

$$\bar\tau_\mu\tau_\nu=\delta_{\mu\nu}+i\sigma_{\mu\nu}\,, \tag{2.15}$$

involving the self-dual spin tensor

$$\sigma_{\mu\nu}=\left[\begin{array}{ccc|c} 0 & \sigma_3 & -\sigma_2 & \sigma_1\\ -\sigma_3 & 0 & \sigma_1 & \sigma_2\\ \sigma_2 & -\sigma_1 & 0 & \sigma_3\\ \hline -\sigma_1 & -\sigma_2 & -\sigma_3 & 0\end{array}\right]. \tag{2.16}$$

Equation (2.14) is the second-order Dirac equation that we are looking for to describe our N-dimensional multiplet of fermions. As in the one-particle case investigated before, the dual wave function $\bar\Phi$ plays a role in the new formalism analogous to the role of $\bar\Psi=\Psi^\dagger\gamma_4$ in the theory of the first-order Dirac equation. The self-adjointness properties of A and B, $(\vec A\Phi)^\dagger=\Phi^\dagger\overleftarrow{A}$, $(\vec B\Phi)^\dagger=\Phi^\dagger\overleftarrow{B}$, make it quite simple to obtain the wave equation for the dual state $\bar\Phi$. We take the adjoint of Eq. (2.10) obtaining $\Phi^\dagger(-\overleftarrow{B}\overleftarrow{A}+m^2)=0$ and then act on the right with $\overleftarrow{B}$. The resulting wave equation is $\Phi^\dagger\overleftarrow{B}(-\overleftarrow{A}\overleftarrow{B}+m^2)=0$, equivalently

$$\bar\Phi\{\overleftarrow{\Pi}\cdot(1+i\sigma)\cdot\overleftarrow{\Pi}+m^2\}=0\,. \tag{2.17}$$

If now Eq. (2.14) for a state Φ_A is multiplied on the left by a dual state $\bar\Phi_B$ obeying Eq. (2.17), then Eq. (2.17) for $\bar\Phi_B$ is multiplied on the right by Φ_A, and the two resulting equations are subtracted; then a conservation law emerges: $\partial_\mu j_\mu=0$, where j_μ is the transition current

$$j\equiv(1/m^2)\bar\Phi_B\{\overleftarrow{\Pi}\cdot(1+i\sigma)+(1+i\sigma)\cdot\vec\Pi\}\Phi_A\,. \tag{2.18}$$

We shall identify the transition current (2.18) with the fermion number.

III. SECOND QUANTIZATION

As indicated in the Introduction, we start out with the conventional path integral treatment of the gauge multiplet based on the linear Dirac equation. Since the path integral technique has been discussed extensively elsewhere,[17] our discussion can be brief, concentrating on the essential new points. We work in terms of the generating functional, G, of the vacuum τ-functions,

$$\langle 0+|T(\Psi_{\rm H}(2)\bar\Psi_{\rm H}(1)\cdots A_\mu(3)_{\rm H}\cdots)|0-\rangle/\langle 0+|0-\rangle\,,$$

where $\Psi_{\rm H}$ is the exact Heisenberg field operator for the array (2.7). The expression for G is

$$\begin{aligned} G&\equiv\int[dA_\mu][d\bar\Psi\,d\Psi]\det(-i\partial\cdot\mathscr{D})\\ &\quad\times\exp\Big(i\int d^4x(\mathscr{L}_{\rm EFF}+J_\mu A_\mu+\bar\Psi\lambda+\bar\lambda\Psi)\Big),\\ \mathscr{L}_{\rm EFF}&=\sum_c\Big(-\frac14F_{\mu\nu C}F_{\mu\nu C}-\frac{1}{2\alpha}(\partial_\mu A_{\mu C})^2\Big)\\ &\quad+\mathscr{L}_{\rm DIRAC}\,, \end{aligned} \tag{3.1}$$

in which

$$\begin{aligned}&\mathscr{L}_{\rm DIRAC}=\bar\Psi(-i\not\Pi-m)\Psi\,,\quad \not\Pi=\Gamma_\mu\Pi_\mu\,,\\ &\Pi_\mu=-i\partial_\mu-gA_\mu\,,\quad A_\mu=A_{\mu C}T_C\,.\end{aligned} \tag{3.2}$$

The matrices T_C are the generators of the symmetry group, and obey the equations of a Lie algebra, $[T_A;T_B]=if_{ABC}T_C$, with structure constants f_{ABC}. The operators $\mathscr{D}_\mu$ are defined by

$$(\mathscr{D}_\mu)_{AB}\equiv-i\partial_\mu\delta_{AB}-g[F_C]_{AB}A_{\mu C}\,, \tag{3.3}$$

where the matrices $[F_C]$ are the matrices of the adjoint representation $[F_C]_{AB}\equiv-if_{CAB}$.

For fixed A_μ the functional integral in Eq. (3.1) over the quark degrees of freedom can be formally performed in closed form:

$$\begin{aligned}&\int[d\bar\Psi\,d\Psi]\exp\Big(\int d^4x(\bar\Psi(\not\Pi-im)\Psi+\bar\Psi\lambda i+i\bar\lambda\Psi)\Big)\\ &\quad=C\det(\not\Pi-im)\exp\Big(-\int d^4x\,d^4y\,i\bar\lambda(x)\\ &\qquad\times\Big\langle x\Big|\frac{1}{\not\Pi-im}\Big|y\Big\rangle i\lambda(y)\Big).\end{aligned} \tag{3.4}$$

Here and subsequently we employ a notation in which propagators are visualized as matrix representatives of abstract operators.[15] We write, for example, $S_F(2,1)=\langle 2|(p^2+m^2)^{-1}|1\rangle$, in which $p_\mu\equiv-i\partial_\mu$, the representation being with respect to a basis of space-time coordinate eigenkets $|1\rangle\equiv|\mathbf r_1,t_1\rangle$ defined through the equations $x_\mu|1\rangle=(x_\mu)_1|1\rangle$ and

$$\langle 2|1\rangle=\delta^4(2,1)\equiv\delta(t_2-t_1)\delta^3(\mathbf r_2-\mathbf r_1).$$

The time coordinate is thus treated on the same footing as x,y,z.

In Eq. (3.4), $\not\Pi$ is defined as in Eq. (3.2) except that now $A_{\mu C}$ represents the abstract operator

$$A_{\mu C}\equiv\int d^4\xi\,|\xi\rangle\langle\xi|A_{\mu C}(\xi)\,, \tag{3.5}$$

in which $A_{\mu C}(\xi)$ is the c-number potential over which the functional integral $\int[dA_\mu]$ in Eq. (3.1) will be performed.

Next we specialize the Grassmann parameters $\lambda,\bar\lambda$ as follows:

$$\lambda=\begin{bmatrix}m\Lambda\\0\end{bmatrix},\quad \bar\lambda=[\bar\Lambda,0]\,. \tag{3.6}$$

In view of the relation (2.7) and a similar one for $\bar\Psi$ we have the identity

$$\bar\Psi\lambda+\bar\lambda\Psi=\bar\Phi\Lambda+\bar\Lambda\Phi\,. \tag{3.7}$$

After this specialization of the parameters of the "Fourier transform" (3.1) it will be possible by functional differentiation to generate only a restricted subset of vacuum τ-functions of the type

$$\langle 0-|T(\Phi_{\rm in}(2)\bar\Phi_{\rm in}(1)\cdots A_\mu(3)_{\rm in}\cdots S)|0-\rangle$$

involving the quark degrees of freedom only through the $\Phi_{\rm in}$'s and $\bar\Phi_{\rm in}$'s. These are, however, exactly the vacuum τ-functions of interest for the scalar formalism.

When the parameters $\lambda,\bar\lambda$ are restricted in this way, the right-hand side of Eq. (3.4) can be transformed into an expression relating it to the second-order Dirac equation.

The relation $(\not{\Pi} - im)^{-1} = (\not{\Pi}\not{\Pi} + m^2)^{-1}(\not{\Pi} + im)$ is used in conjunction with the expressions (2.9) for the Γ_μ to obtain

$$\frac{1}{\not{\Pi} - im} = \begin{bmatrix} g_+ im & g_+ i\bar{\tau}\cdot\Pi \\ -g_- i\tau\cdot\Pi & g_- im \end{bmatrix}, \tag{3.8}$$

in which

$$g_+ = (\bar{\tau}\cdot\Pi\tau\cdot\Pi + m^2)^{-1}$$

and

$$g_- = (\tau\cdot\Pi\bar{\tau}\cdot\Pi + m^2)^{-1}.$$

When the expectation value

$$\int d^4x\, d^4y\, \bar{\lambda}(x)\langle x|(\not{\Pi} - im)^{-1}|y\rangle\lambda(y)$$

in the exponential of Eq. (3.4) is computed using Eq. (3.8) and using the restricted forms (3.6) of $\lambda,\bar{\lambda}$, the exponential goes over into an expression involving the propagator of the scalar formalism:

$$\int d^4x\, d^4y\, \bar{\lambda}(x)\langle x|(\not{\Pi} - im)^{-1}|y\rangle\lambda(y)$$
$$= \int d^4x\, d^4y\, \bar{\Lambda}(x)$$
$$\times \langle x|im^2(\Pi\cdot(1+i\sigma)\cdot\Pi + m^2)^{-1}|y\rangle\Lambda(y). \tag{3.9}$$

This relation uses the identity $\bar{\tau}\cdot\Pi\tau\cdot\Pi = \Pi\cdot(1+i\sigma)\cdot\Pi$ [see Eq. (2.15)].

The next step is to rewrite the determinant $\det(\not{\Pi} - im)$ on the right-hand side of Eq. (3.4) in terms relating it to the scalar formalism. This is achieved through use of the matrix identity

$$\det\begin{bmatrix} \lambda & C \\ D & \lambda \end{bmatrix} = \det(\lambda^2 - CD), \tag{3.10}$$

in which λ signifies a square matrix proportional to the identity matrix, and C and D are arbitrary square matrices having the same dimension as λ. Formal application of the identity (3.10) to the operator

$$\not{\Pi} - im = \begin{bmatrix} -im & i\bar{\tau}\cdot\Pi \\ -i\tau\cdot\Pi & -im \end{bmatrix}$$

yields the identity

$$\det(\not{\Pi} - im) = \det(-(\bar{\tau}\cdot\Pi\tau\cdot\Pi + m^2))$$
$$= \det(-(\Pi\cdot(1+i\sigma)\cdot\Pi + m^2)). \tag{3.11}$$

When the results (3.9) and (3.11) are incorporated in Eq. (3.4), and Eq. (3.4) is incorporated in Eq. (3.1), we obtain the generating functional G in the form

$$G \equiv \int [dA_\mu]\det(-i\partial\cdot\mathscr{D})\exp\left(i\int d^4x(\mathscr{L}_{\rm GB} + J_\mu A_\mu)\right)C\det(-(\Pi\cdot(1+i\sigma)\cdot\Pi + m^2))$$
$$\times\exp\left(\int d^4x\, d^4y\, i\bar{\Lambda}(x)\langle x|(-im^2)(\Pi\cdot(1+i\sigma)\cdot\Pi + m^2)^{-1}|y\rangle\, i\Lambda(y)\right). \tag{3.12}$$

Here $\mathscr{L}_{\rm GB}$ signifies the Lagrangian density for the gauge degrees of freedom.

Performing the functional integral $\int[dA_\mu]$ in Eq. (3.12) gives

$$G = C\exp\Bigg\{-i\int d^4x\,\Omega_{\nu C}(-i\overleftarrow{\partial}_\mu)g[F_C]_{AB}\Omega_{\mu A}\Omega_{\nu B} + i\int d^4x\frac{g^2}{4}[F_C]_{AB}[F_C]_{DE}\Omega_{\mu A}\Omega_{\nu B}\Omega_{\mu D}\Omega_{\nu E}$$
$$-\sum_{n=1}^{\infty}\frac{1}{n}\operatorname{tr}(p^{-2}p_\mu g[F_C]\Omega_{\mu C})^n - \sum_{n=1}^{\infty}\frac{1}{n}\operatorname{tr}((p^2 + m_0{}^2)^{-1}\mathscr{V})^n$$
$$+\int d^4x\, d^4y\, i\bar{\Lambda}(x)\langle x|(-im_0{}^2)(\Pi\cdot(1+i\sigma)\cdot\Pi + m_0{}^2)^{-1}|y\rangle\, i\Lambda(y)\Bigg\}$$
$$\times\exp\left\{\frac{1}{2}\int d^4x_2\, d^4x_1\, iJ_{\alpha K}(2)(-i)D_F(2,1)_{\alpha\beta}\, iJ_{\beta K}(1)\right\}. \tag{3.13}$$

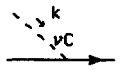
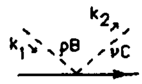
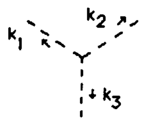
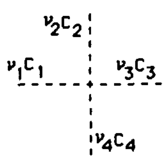
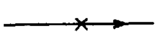
The notations here are

$$\mathscr{V} \equiv g(p\cdot(1+i\sigma)\cdot\Omega + \Omega\cdot(1+i\sigma)\cdot p)$$
$$-g^2\Omega\cdot(1+i\sigma)\cdot\Omega,$$
$$\Pi_\mu = -i\partial_\mu - g\Omega_\mu, \quad \Omega_\mu = \Omega_{\mu C}T_C,$$
$$\Omega_{\mu C} \equiv \int d^4\xi\, |\xi\rangle\langle\xi|\Omega_{\mu C}(\xi),$$
$$\Omega_{\mu C}(\xi) \equiv \delta/\delta iJ_{\mu C}(\xi),$$
$$-iD_F(2,1) \equiv \left\langle 2\left|\frac{-i}{k^2}\left[\delta_{\mu\nu} - (1-\alpha)\frac{k_\mu k_\nu}{k^2}\right]\right|1\right\rangle,$$
$$k_\mu \equiv -i\partial_\mu.$$

The representation (3.13) provides a convenient means of generating compact analytic representations of Feynman integrals. The two infinite series in Eq. (3.13) contain the closed loop diagrams for "ghosts" and quarks, and arise from an expansion of determinants using the matrix identity $\det(M) = \exp(\operatorname{tr}(\ln(M)))$.

Table I provides a summary of the Feynman rules obtained by the perturbation expansion of Eq. (3.13). As indicated in the Introduction, the rules of Table I generalize earlier results obtained for the Abelian case of quantum electrodynamics.[4,5,14] In order to avoid dealing with complicated mass counterterms we have worked above entirely in terms of the bare mass m_0. Accordingly, Feynman integrals obtained by expanding Eq. (3.13) would come out at first in terms of the bare mass. Conversion to the physical mass is carried out quite simply by replacing each occurrence of the propagator $1/(p^2 + m_0{}^2)$ by $1/(p^2 + m^2 - \delta m^2)$ and then

TABLE I. Scalar formalism for gauge theories. Feynman rules for the calculation of τ-functions.[a]

Graph	Name	Factor
	quark line	$1/(p^2+m^2)$
	internal gluon line	$\int \frac{d^4k}{(2\pi)^4}\frac{-i}{k^2}\left[\delta_{\mu\nu}-(1-\alpha)\frac{k_\mu k_\nu}{k^2}\right]\delta_{AB}$
	ghost line	$-1/p^2$
	simple quark gluon vertex	$g(p\cdot(1+i\sigma)T_C e^{ik\cdot x}+e^{ik\cdot x}T_C(1+i\sigma)\cdot p)_\nu$
	double quark gluon vertex	$-g^2[\delta_{\nu\rho}\{T_C;T_B\}+i\sigma_{\nu\rho}[T_C;T_B]]\,e^{ik_1\cdot x-ik_2\cdot x}$
	three gluon vertex	$gf_{C_3C_2C_1}[\delta_{\nu_2\nu_1}(k_2-k_1)_{\nu_3}+\delta_{\nu_3\nu_2}(k_3-k_2)_{\nu_1}+\delta_{\nu_1\nu_3}(k_1-k_3)_{\nu_2}]$
	four gluon vertex	$-ig^2[f_{EC_4C_2}f_{EC_1C_3}(\delta_{\nu_1\nu_4}\delta_{\nu_3\nu_2}-\delta_{\nu_1\nu_2}\delta_{\nu_3\nu_4})$ $+f_{EC_3C_2}f_{EC_4C_1}(\delta_{\nu_4\nu_3}\delta_{\nu_1\nu_2}-\delta_{\nu_4\nu_2}\delta_{\nu_1\nu_3})$ $+f_{EC_3C_4}f_{EC_1C_2}(\delta_{\nu_1\nu_3}\delta_{\nu_2\nu_4}-\delta_{\nu_1\nu_4}\delta_{\nu_2\nu_3})]$
	ghost gluon vertex	$gF_C\,p_\nu e^{ik\cdot x}$
	mass counterterm	δm^2

[a] The rules give $\tau/(-im_0{}^2)^{\overline{Q}}$, where $\overline{Q}$ is the number of external quark lines, and

$\tau \equiv \langle 0-|T(\Phi(2)_{\rm in}\overline{\Phi}(1)_{\rm in}\cdots A_\mu(3)\cdots S)|0-\rangle/\langle 0-|S|0-\rangle$.

In addition to the rules of Table I, we must insert a factor (-1) for each closed fermion loop, including ghost loops. Both quark and ghost particles are treated using an abstract operator notation in which $p\equiv -i\partial$ [see the discussion after Eq. (3.4)]. For a quark loop a trace over space-time, spin, and internal degrees of freedom is required. For a ghost loop a trace over space-time and internal degrees of freedom is required.

expanding in powers of δm^2. This is the origin of the δm^2 vertex in Table I.

As a final step in our treatment of gauge field theory by the scalar formalism, we note that the integrand of Eq. (3.12) has a representation in terms of a path integral over Grassmann fields $\Phi,\overline{\Phi}$. We thus arrive at a representation of the generating functional in the form

$$G=\int [dA_\mu][d\overline{\Phi}\,d\Phi]\det(-i\partial\cdot\mathscr{D})\exp\left(i\int d^4x\,\mathscr{L}\right), \tag{3.14}$$

$$\mathscr{L}\equiv\sum_C\left(-\frac{1}{4}F_{\mu\nu C}F_{\mu\nu C}-\frac{1}{2\alpha}(\partial_\mu A_{\mu C})^2\right)$$
$$-(1/m_0{}^2)\overline{\Phi}(\overleftarrow{\Pi}\cdot(1+i\sigma)\cdot\overrightarrow{\Pi}+m_0{}^2)\Phi,$$
$$\Pi_\mu=-i\partial_\mu-gA_\mu,\quad A_\mu=A_{\mu C}T_C,$$

in which an effective "Lagrangian density"

$$\mathscr{L}_{\rm DIRAC}\equiv-(1/m_0{}^2)\overline{\Phi}(\overleftarrow{\Pi}\cdot(1+i\sigma)\cdot\overrightarrow{\Pi}+m_0{}^2)\Phi \tag{3.15}$$

of the second-order Dirac equation (2.4) appears. The effective density (3.15) has the same general structure as the Lagrangian density introduced by Brown[4] for the Abelian case of quantum electrodynamics.

ACKNOWLEDGMENTS

Part of this work was carried out in the summer of 1985 while the author was a Visiting Scientist in the Physics Department at Cornell University. I am grateful to the Physics Department at Cornell University for their hospitality.

This research was supported by the National Science Foundation under Grant No. PHY-8415543.

APPENDIX: TREATMENT OF THE FERMION DEGREES OF FREEDOM BY THE CANONICAL FORMALISM AND THE DYSON–WICK EXPANSIONS

The equivalence of the new and the old formalism may be established treating the fermion degrees of freedom by the canonical formalism and the Dyson–Wick expansions,[18] along the lines of Ref. 14. Since the gauge field is most simply quantized by path integral techniques, we shall first go over to an intermediate interaction picture in which the virtues of both approaches may be exploited.

We once again begin with the conventional gauge field theory of the multiplet. Therefore we start by second quantizing the fermion field Ψ of the linear Dirac equation (2.1) in the usual way. Then second-quantized fields Φ and $\overline{\Phi}$ are defined through the prescription (2.7). The desired Feynman rules are obtained by then rewriting all in terms of the arrays of Pauli spinors Φ and $\overline{\Phi}$, and using the Dyson–Wick expansions.

We start out in the Coulomb gauge, since the canonical formalism is particularly straightforward to implement in this gauge. We shall use the shorthand notation q, p to refer to the canonical variables of the gauge field. The total Hamiltonian of the system is written in a usual way as $H = H_{\mathrm{GB}}(q,p) + H_Q(\Psi,\overline{\Psi}) + H_I(q,p,\Psi,\overline{\Psi})$. We assume all variables $q, p, \Psi, \overline{\Psi}$ to be in the Schrödinger picture, defined to coincide with the Heisenberg picture at a time t_i so far in the remote past that the interaction Hamiltonian H_I becomes ineffective and can be neglected.

The intermediate interaction picture referred to above is defined by separating out a factor $\exp(-iH_Q(t-t_i))$ from the exact Schrödinger picture time evolution operator $\exp(-iH(t-t_i))$. Note that the interaction picture defined in this way is an incomplete interaction picture in that the factor separated out, $\exp(-iH_Q(t-t_i))$, does not incorporate the free Hamiltonian of the gauge field. For this reason all dynamical variables referring to the gauge degrees of freedom are left invariant, and remain in the Schrödinger picture. The effective interaction picture Hamiltonian is $H_{\mathrm{GB}}(q,p) + H_{\mathrm{INT}}(q,p,\Psi_{\mathrm{in}},\overline{\Psi}_{\mathrm{in}})$. That this is the full Schrödinger picture Hamiltonian in the gauge fields makes it possible to treat the gauge boson degrees of freedom in an expression

$$\tau(2,1) = \langle 0- | T\Big(\Phi_{\mathrm{in}}(2)\overline{\Phi}_{\mathrm{in}}(1)\exp\Big[-i\int_{-\infty}^{\infty} dt\,(H_{\mathrm{GB}} + H_{\mathrm{INT}})\Big]\Big)|0-\rangle(\langle 0+|0-\rangle)^{-1}$$

for a vacuum τ-function in the usual way by use of the path integral technique. The resulting expression for the τ-function is

$$\begin{aligned}
&\tau(2,1) = N/D\,,\\
&N = \int [dA_\mu]\det(-i\partial\cdot\mathscr{D})_Q \langle 0-|T\Big(\Phi_{\mathrm{in}}(2)\overline{\Phi}_{\mathrm{in}}(1)\exp\Big(i\int d^4x\,\mathscr{L}_{\mathrm{EFF}}\Big)\Big)|0-\rangle_Q\,,\\
&D = \int [dA_\mu]\det(-i\partial\cdot\mathscr{D})_Q \langle 0-|T\Big(\exp\Big(i\int d^4x\,\mathscr{L}_{\mathrm{EFF}}\Big)\Big)|0-\rangle_Q\,,\\
&\mathscr{L}_{\mathrm{EFF}} = \sum_C\Big(-\frac{1}{4}F_{\mu\nu C}F_{\mu\nu C} - \frac{1}{2\alpha}(\partial_\mu A_{\mu C})^2 + j_{\mu C}A_{\mu C}\Big)\,.
\end{aligned} \tag{A1}$$

The Heisenberg vacuum in the remote past is here assumed to factorize into an outer product of a vacuum for the gauge field and a vacuum for the quark field $|0-\rangle = |0-\rangle_{\mathrm{GB}} \otimes |0-\rangle_Q$. Only $|0-\rangle_Q$ survives in Eq. (A1).

The identity

$$j_C = (g/m^2)\overline{\Phi}_{\mathrm{in}}(\overleftarrow{p}\cdot(1+i\sigma) + (1+i\sigma)\cdot\vec{p})T_C\Phi_{\mathrm{in}} \tag{A2}$$

allows us to express the τ-function (A1) entirely in terms of the in fields Φ_{in} and $\overline{\Phi}_{\mathrm{in}}$. Feynman rules for the τ-function may now be obtained by making a perturbation expansion and using the Dyson–Wick procedure, along the lines of Ref. 14. Double gluon–quark vertices, or seagulls, characteristic of a theory of scalar particles arise in the same way as in the quantum electrodynamics treated earlier in Ref. 14.

[1]We use natural units defined by $\hbar = c = 1$. Four vectors have an imaginary time component, e.g., $x_\mu = (\mathbf{r}, it)$. Accordingly, the Lorentz metric is the simple $\delta_{\mu\nu}$. Derivative operators acting to the left signify *minus* differentiation of the objects on the left, e.g., $\overline{\Phi}\overleftarrow{\partial}_\mu \equiv -\partial\overline{\Phi}/\partial x_\mu$.

[2]O. Laporte and G. E. Uhlenbeck, Phys. Rev. **37**, 1380 (1931).
[3]R. P. Feynman and M. Gell-Mann, Phys. Rev. **109**, 193 (1958).
[4]L. M. Brown, Phys. Rev. **111**, 957 (1958).
[5]M. Tonin, Nuovo Cimento **14**, 1108 (1959).
[6]W. R. Theis, Fortsch. Phys. **7**, 559 (1959).
[7]H. Pietschmann, Acta Phys. Austriaca **14**, 63 (1961).
[8]L. M. Brown, "Two-component fermion theory," in *Lectures in Theoretical Physics*, Vol. IV (Interscience, New York, 1962).
[9]L. M. Brown, "Quantum electrodynamics at high energy," in *Topics in Theoretical Physics, Proceedings of the Liperi Summer School in Theoretical Physics*, 1967, edited by C. Cronstrom (Gordon and Breach, New York, 1969), p. 113.
[10]P. R. Auvil and L. M. Brown, Am. J. Phys. **46**, 679 (1978).
[11]L. C. Hostler, J. Math. Phys. **23**, 1179 (1982).
[12]L. C. Hostler, J. Math. Phys. **24**, 2366 (1983).
[13]L. C. Hostler, J. Math. Phys. **26**, 124 (1985).
[14]L. C. Hostler, J. Math. Phys. **26**, 1348 (1985).
[15]J. Schwinger, Proc. Nat. Acad. Sci. **37**, 455 (1951).
[16]Reference 11, Eq. (2.10).
[17]The following are references on gauge theories and the path integral method in quantum field theory: L. D. Faddeev and A. A. Slavnov, *Gauge Fields, Introduction to Quantum Theory* (Benjamin Cummings, Reading, MA, 1980); P. H. Frampton, *Lectures on Gauge Field Theories*, lectures given in graduate seminar at UCLA, Spring quarter, 1977; T. D. Lee, *Particle Physics and Introduction to Field Theory* (Harwood Academic, New York, 1981); C. Itzykson and J.-B. Zuber, *Quantum Field Theory* (McGraw–Hill, New York, 1980).
[18]For the canonical formalism and the Dyson–Wick treatment, see, for example, J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw–Hill, New York, 1965).

Spectra of supersymmetric O(*N*) sigma models in 0 +1 dimension

B. Sathiapalan
Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104

(Received 28 February 1986; accepted for publication 30 April 1986)

The spectra of the supersymmetric O(N) nonlinear sigma models in 0 + 1 space-time dimension are computed exactly for any N, using group theoretical methods. The allowed representations are O(N) analogs of the Wigner d functions.

I. INTRODUCTION

Recently, it was found[1] that the supersymmetric O(3) sigma model[2] in 0 + 1 dimension with a Wess–Zumino term[3] has an interesting spectrum: both the ground and first excited levels have unequal numbers of Bose and Fermi states. This is in contrast to what is normally found in the spectra of supersymmetric theories where the ground state alone is unpaired (if supersymmetry is not broken by the vacuum) and all other levels have equal numbers of Bose and Fermi states. This then raises the exciting possibility that the observed asymmetry between bosons and fermions need not have anything to do with the breaking of supersymmetry. In order to explore this line of thought, however, one needs more examples of theories where such atypical representations of supersymmetry arise.

It is tempting to first look for other examples in 0 + 1 space-time dimensions since they often can be solved exactly. Besides being easy to solve, (0 + 1)-dimensional theories often can be of direct relevance to higher-dimensional field theories: if a field theory has soliton solutions then the dynamics of the collective coordinates of the solitons will be described by a (0 + 1)-dimensional theory, the spectrum of which is in fact the spectrum of soliton states in the original field theory. However, the list of exactly soluble theories that have any resemblance to higher-dimensional field theories is surprisingly short. It is the purpose of this paper to remedy this situation somewhat by extending this list. We compute, using group theoretical methods, the spectrum of the supersymmetric O(N) nonlinear σ-models (in 0 + 1 dimensions) for arbitrary N. We do this both for the case of $N = 1$ and the case of $N = \frac{1}{2}$ supersymmetry. The O(3) σ model was solved in Ref. 4. Certain aspects of the O(N) model were studied in Ref. 5. In Ref. 4 the solutions were found to be the Wigner d-functions that describe a symmetric top.[5] In computing the spectrum of the O(N) model we are led to extend this concept of the d function to representations of the O(N) group. In this paper, we have not, however, tried to explicitly construct these representatives as a set of functions defined on S^{N-1}.

II. DIRAC QUANTIZATION OF THE O(*N*) SUPERSYMMETRIC σ MODEL

The O(N) supersymmetric σ model is defined by he Lagrangian

$$L = \left[\tfrac{1}{2}(\partial_t n^a)^2 + (i/2)\bar{\psi}\,\partial\psi + \tfrac{1}{8}(\bar{\psi}^a\psi^a)\right] \tag{1a}$$

and the constraints

$$n^a n^a = 1, \quad n^a\psi^a = 0 \quad \bar{\psi}^a = \psi^T\gamma^0 = \psi^T\sigma_2\,,$$

$$\text{where } a = 1,2,\ldots,N. \tag{1b}$$

Here n^a is an N-component vector and ψ^a is a two-component Majorana spinor and also an N vector. This defines the $M = 1$ supersymmetric theory and to get the $M = \frac{1}{2}$ supersymmetry theory we set the lower component of the two-spinor equal to zero. We now proceed to quantize this system using Dirac's procedure for constrained Hamiltonian systems generalized to include fermions.[6] We consider the $M = 1$ case. The $M = \frac{1}{2}$ will be taken up in the end.

The Hamiltonian is given by

$$H = (P^2/2) - \tfrac{1}{8}(\bar{\psi}\psi)^2\,. \tag{2}$$

In addition to the two constraints given earlier, there is a secondary constraint $n \cdot P = 0$ and an additional constraint coming from the Majorana property of the fermion: $\pi^{\alpha a} + (i/2)\psi^{\alpha a} = 0$, where $\pi^{\alpha a}$ is the momentum conjugate to $\psi^{\alpha a}$. One can now construct Dirac brackets in ther standard way:

$$\begin{aligned}
\{n^a, p^b\}^* &= \delta^{ab} - n^a n^b\,,\\
\{p^a, p^b\}^* &= n^b p^a - n^a p^b + (i/2)[\psi^a, \psi^b]\,,\\
\{\psi^{\alpha a}, p^b\}^* &= -n^a\psi^{\alpha b}\,,\\
\{\psi^{\alpha a}, \psi^{b\beta}\} &= -i\delta^{\alpha\beta}(\delta_{ab} - n_a n_b)
\end{aligned} \tag{3}$$

($\alpha = 1,2$ is the "Dirac" index).

Note that we have written the product $\psi^a\psi^b$ as a commutator. This automatically guarantees that there are no ordering ambiguities on quantizing. The commutation and anticommutation relations are obtained by multiplying the Dirac brackets by the factor "i." Using Eq. (3) one can verify that

$$\{Q^\alpha, Q^\beta\} = \delta^{\alpha\beta}H = \frac{\delta^{\alpha\beta}}{2}\left[p^2 - \frac{(\psi\psi)^2}{4} + \frac{(N-1)^2}{4}\right], \tag{4}$$

where $Q^\alpha = \psi^\alpha \cdot p$ is the supercharge operator. This differs from (2) by a constant reflecting the ordering ambiguities in the clasical expression for the Hamiltonian. One can further verify that

$$H = \tfrac{1}{2}J^2\,, \tag{5}$$

where $J^{ab} = n^{[a}p^{b]} - (i/2)[\psi^a, \psi^b]$ are the generators of O(N). The ordinary O(N) nonlinear sigma model merely describes a rigid rotor in N dimensions and the fact that the Hamiltonian is equal to the quadratic Casimir of O(N) is not surprising. What is perhaps nontrivial is that just as was found in the O(3) model, this equality generalizes to the supersymmetric case as well for all the O(N) groups. The main differences between the supersymmetric and nonsupersymmetric cases is that in the former there are certain constraints on what eigenvalues of J^2 occur in the spectrum and their multiplicity. We first describe the spectrum of the O(3) σ model solved in Ref. (4) and we then generalize to O(N).

III. SPECTRUM AND REPRESENTATIONS

The constraints on eigenvalues of J^2 are determined as follows: Consider the quantity

$$\mathbf{n}\cdot\mathbf{J}=\tfrac{1}{2}\varepsilon^{abc}n^a J^{bc}=-(i/4)\varepsilon^{abc}n^a\psi^{[b}\psi^{c]}\ .$$

This is the radial component of the total angular momentum. Let us choose a basis where $n=(0,0,1)$. Then

$$n\cdot J=J_3=-(i/2)[\psi^1,\psi^2]\ . \tag{6}$$

If we define $\psi_\alpha=(\psi^1_\alpha+i\psi^2_\alpha)/\sqrt{2}$, so that $\{\psi_\alpha,\psi_\beta\}=\{\psi^*_\alpha,\psi^*_\beta\}=0$ and $\{\psi_\alpha,\psi^*_\beta\}=\delta_{\alpha\beta}$ we get $n\cdot J=\frac{1}{2}[\psi,\psi^*]$. Now let us define the state $|0>$ by $\psi_\alpha|0\rangle=0$ (α is the Dirac index). We can then construct the three states $\psi^*_1|0\rangle$, $\psi^*_2|0\rangle$, and $\psi^*_1\psi^*_2|0\rangle$. One then finds that the states $|0\rangle$, $\psi^*_1|0\rangle$, $\psi^*_2|0\rangle$, and $\psi^*_1\psi^*_2|0\rangle$ have $n\cdot J$ equal to $+1,0,0$, and -1, respectively. Clearly, then for the bosonic states $|0\rangle$ and $\psi^*_1\psi^*_2|0\rangle$ the eigenvalues of J^2 are $j(j+1)$ and $j\geqslant 1$ since the value 0 is not allowed, whereas for the remaining two states all integer values of $j\geqslant 0$ are allowed (Fig. 1). Thus there are two fermionic zero energy states, which shows that the index is -2 and supersymmetry is unbroken in the corresponding higher-dimensional field theories. It is also known[7] that the index is equal to the Euler characteristic of the manifold, which is 2 (see Ref. 8) for S^2. The wave function describing the states with $n\cdot J>0$ are generalizations of the usual spherical harmonics and are the Wigner d functions. Thus two quantum numbers are needed to specify a representation J^2 and $n\cdot J$. This feature has a generalization in the O(N) case.

Let us turn to the O(N) σ model. We shall consider in detail the case where N is odd and see that the results can easily be extrapolated to even N also. The generalization of Eq. (6), where $\mathbf{n}=(0,0,\ldots,1)$, is

$$\tfrac{1}{2}\varepsilon^{abcd\cdots}n^a J^{bc}\sim S^{ij}=-(i/2)\psi^{[i}\psi^{j]}\ , \tag{7}$$

where i,j run from 1 to $(n-1)$ and are tangential to S^{N-1}. This forms an O($n-1$) "spin" subgroup of the total angular momentum J. We can find out what the constraints on allowed representations of S^{ij} are by studying the eigenvalues of the Cartan subalgebra of S^{ij}, which we take to be S^{12}, $S^{34},\ldots,S^{2p-1\,2p}$, where $N=2p+1$. The operators

$$\psi^{\rm I}_\alpha=\frac{\psi^1_\alpha+i\psi^2_\alpha}{\sqrt{2}},\quad \psi^{\rm II}_\alpha=\frac{\psi^3_\alpha+i\psi^4_\alpha}{\sqrt{2}},\ldots,\psi^P_\alpha=\frac{\psi^{2p-1}_\alpha+i\psi^{2p}_\alpha}{\sqrt{2}}$$

are constructed and the state $|0\rangle$ defined as the "Fock vacuum" for these fermionic operators. The states $\psi^{*J}_\alpha\psi^{*K}_\alpha\cdots|0\rangle$ form the various sectors with different "fermion number." One can calculate the eigenvalues of the generators of the Cartan subalgebra, transform to the more convenient Chevalley basis,[9] and see what representations of O($N-1$) are obtained. If $h_1,h_2,\ldots,h_p$ denote the eigenvalues of $S^{12},S^{34},\ldots,S^{2p-1\,2p}$ then, in the Chevalle basis, representations are labeled by the sequence of numbers $h_{\alpha i}=h_i-h_{i+1},\ldots,h_{\alpha_{p-1}}=h_{p-1}-h_p$, where $h_{\alpha_p}=h_{p-1}+h_p\cdot\alpha_i$ corresponds to the roots in the Dynkin diagram for O($2p$) given in Fig. 2.

In Table I we have worked it out in detail for the O(5) σ model. In one of the columns, values of a U(1) charge that act on the Dirac index are indicated. This is helpful in assigning states to representations. The representations turn out to be (1,3) + (3,1) + (1,1) + (1,1) [of O(4)] for the Bose states and (2,2) + (2,2) for the Fermi states. In the O(7) model the representations [of the O(6) "spin" subgroup] turn out to be $\underline{10}+\underline{10}+\underline{6}+\underline{6}$ and $\underline{15}+\underline{15}+\underline{1}+\underline{1}$, respectively. The pattern is obvious—they are the n-index antisymmetric representations of O($2p$) with n ranging from 0 to $2p$, alternately even and odd n corresponding to Bose and Fermi states. Thus one finds that only those representations of O(N) are allowed that transform as listed above under its O($n-1$) spin subgroup.[10]

Let us concentrate for a moment on the O(5) model. One might try to find a representation of O(5) that transforms under an O(4) spin subgroup as some subset of (3,1) + (1,3) + (1,1) + (1,1) for the Bose states. But by the same token this representation of O(5) would also have to transform under O(4) as a subset of (2,2) + (2,2), since by supersymmetry the Fermi states have to transform under O(5) in exactly the same way as the Bose states. But this is clearly impossible since the two sets have no representation in common. The resolution of this problem lies in the fact that the "orbital" part L^{ij} of the O($N-1$) subgroup has been ignored. If we let $n=(0,0,\ldots,1)+(x_1,x_2,\ldots,x_{N-1},0)$ for points infinitesimally away from the "north" pole, the $\mathbf{x}$ transforms as an $\underline{N-1}$ of O($N-1$) and the wave function $\psi(x)$ transforms nontrivially under L^{ij}. [Note that $L^{ij}\psi(\mathbf{x}=0)=0$ because, either ψ does not depend on $\mathbf{x}$, or if it does depend on $\mathbf{x}$, ψ must necessarily vanish at $x=0$ for it to be well defined—this is analogous to the fact that on a two-sphere wave functions that have a nontrivial ϕ dependence vanish at the north pole.] Thus the wave function has the structure $\Sigma_a x^i x^j\cdots x^k r_a{}^{mn\cdots}$, where i,j are indices that transform under L and m,n transform under S. We already know what representations r_a of S are allowed. Thus we can conclude that only those representations of O(N) that trans-

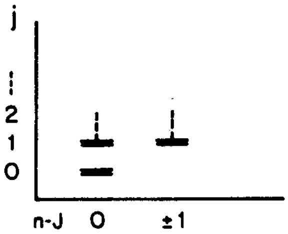


FIG. 1. The spectrum of the O(3) supersymmetric σ model in 0 + 1 dimensions. The energy is given by $\frac{1}{2}j(j+1)$.

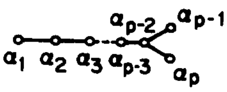


FIG. 2. Dynkin diagram for O($2p$). The α_i are the simple roots.

TABLE I. Transformation properties of the allowed O(5) representations under the O(4) spin subgroup. The U(1) charge is associated with the Dirac index.

State	h_1	h_2	U(1)	h_{α_1}	h_{α_2}	No. of states	O(4)Rep.
$\|0\rangle$	1	1	0	0	2	1	$(3,1)\oplus(1,3)$
$\psi_\alpha^{J+}\|0\rangle$	0	1	$\pm\frac{1}{2}$	± 1	1	4	(2,2)
	1	0					
$\psi_\alpha^{J+}\psi_\beta^{K+}\|0\rangle$	0	0	± 1	0	0	2	(1,1)
	0	0	0	0	0	2	$(3,1)\oplus(1,3)$
	± 1	∓ 1	0	± 2	0	2	$(3,1)\oplus(1,3)$
$\psi_\alpha^{J+}\psi_\beta^{K+}\psi_\gamma^{L+}\|0\rangle$	-1	0	$\pm\frac{1}{2}$	± 1	-1	4	(2,2)
	0	-1					
$\psi_1^{I\,+}\psi_2^{I\,+}\psi_1^{I\,+}\psi_2^{II\,+}\|o\rangle$	-1	-1	0	0	-2	1	$(3,1)\oplus(1,3)$

form under $O(N-1)$ as $\Sigma_a x^{i_1}x^{i_2}\cdots x^{i_k} r_a{}^{m_1 m_2\cdots}$ are allowed. In fact since S_{ab} is a tensor under J we have $[J,S^2]=0$. This means that a given $O(N)$ representation can contain precisely one representation of spin-0 $(N-1)$, i.e., all the r_a have to be the *same*. We can see why, under the spin subgroup O(4) of O(5), the Bose and Fermi states can transform differently and yet transform the same way under the full O(4) subgroup. All that needs to be done is to introduce an extra factor of x^i in one of them. The origin of this extra orbital index is also clear — the supersymmetry charge $Q=\psi^i P_i$ has exactly this extra orbital index on P.

There is also another criterion that has to be satisfied. As can be verified by explicit calculation,

$$J^2=L^2+S^2. \tag{8}$$

The $L\cdot S$ term turns out to vanish, which reflects the fact that $L\cdot S$ has only the components L^{ij} that lie entirely in the $O(N-1)$ subgroup and act on the tangent space of S^{N-1} and we have already seen that these operators acting on a wave function give zero. Thus we can conclude that[11]

$$J^2>S^2. \tag{9}$$

From (8) we conclude that invariance under J implies invariance under L and S separately. Thus a term of the form $x^i f^i$ cannot be part of an $O(N)$ singlet. This is reasonable since the $O(N)$ generalization of $x^i f^i$ would be a term of the form $n^a\psi^a$, which we know vanishes. One can now apply these rules to see what the ground state looks like in these theories. If $J^2=0$, clearly $S^2=0$ and therefore in the O(5) model there can be two Bose states with $J^2=0$ and no Fermi states. Thus the supersymmetry index is 2 and supersymmetry is unbroken in the corresponding field theories. The same

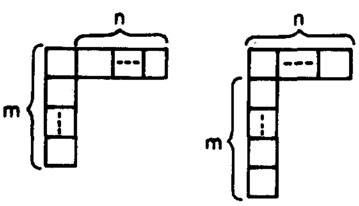


FIG. 3. Young tableaux for representations of the form $x^{(i_1}\cdots x^{i_n)}\, r^{[j_1\cdots j_m]}$.

holds true for the O(7) model and for all the $O(2p+1)$ models. This is consistent with the fact that the supersymmetry index is equal to the Euler characteristic for these models, which is equal to 2 for S^{2p}. In the case of $O(2p)$, since by analogy the representations are n-index antisymmetric tensors with n ranging from O to $2p-1$, it is clear that once again there are two states that have $S^2=0$, but this time one of them is Bose (say the tensor with no index) and one is Fermi (say the one with $2p-1$ indices). Thus the supersymmetry index is zero but the vacuum is still invariant under supersymmetry in the corresponding higher-dimensional field theories. We also note that it is consistent with the fact that the Euler characteristic of S^{2p-1} is zero.

We now proceed to identify the higher representations of $O(N)$ that are allowed. The $O(N-1)$ representations into which it has to decompose were of the form

$$x^{(i}x^j x^k\cdots)\, r_a{}^{[m,n,\cdots]}.$$

This clearly corresponds to the Young tableaux shown in Fig. 3.

The only representations of $O(N)$ that reduce to the above representations of $O(N-1)$ and *nothing else* are the ones shown in Fig. 4, for some p and q. Note that $p+1<N$ for $O(N)$. The value of J^2 for the representation in Fig. 4 is[12] $(q+1)(q+N-1)+(p-1)(N-p-1)$. The value of S^2 for the group $O(N-1)$ is $m(N-1-m)$. Thus as long as $m<p<N/2$, Eq. (9) is satisfied. Since m ranges from 0 to $N-1$ as long as $p<N/2$ we have an allowed representation for any value of q. This then is a complete description of the spectrum of the $O(N)$ σ model. As an example, let us apply these considerations to the O(5) model. The representation of the O(4) spin subgroup were $(3,1)\oplus(1,3)\oplus(1,1)\oplus(1,1)$ and $(2,2)\oplus(2,2)$. These correspond to the Young tableaux shown in Fig. 5(a). The representations of O(5) that are allowed are shown in Fig. 5(b). One has to decide what the multiplicities are. At the ground level there

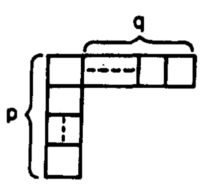


FIG. 4. Young tableaux for the representations of $O(N)$ that reduce to those shown in Fig. 3. Note that $p=1<N$.

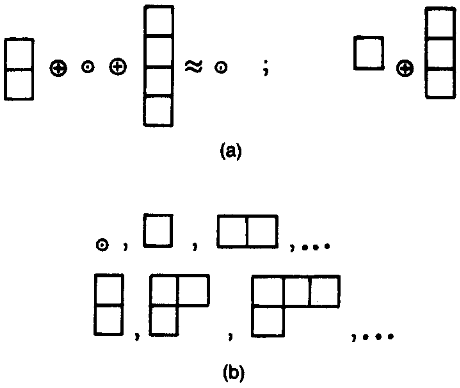


FIG. 5. (a) Representations of the O(4) "spin" group in the Bose sector (3,1) + (1,3) + (1,1) + (1,1) and the Fermi sector (2,2) + (2,2). (b) Representations of O(5) that reduce to those in (a).

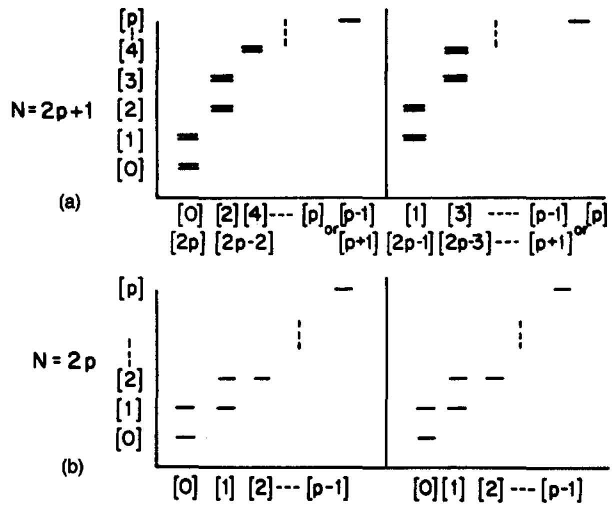


FIG. 7. Spectrum of the O(N) supersymmetric σ model in 0 + 1 dimension. Only the lowest members of each spin series is shown. Here [r] denotes the n index antisymmetric representation of O($N-1$) on the x axis and of O(N) on the y axis. In (a) the duality [n] $\sim$ [$N-n-1$] leads to the observed doubling of states. In (b) this has the effect that all values of n from 0 to $p-1$ are allowed both for the Fermi and Bose series.

are two states with $J^2 = 0$ corresponding to the two singlets in Fig. 5(a). The next representation is $\underline{5}$, which breaks up into a $\underline{4}$ [= (2,2)] + $\underline{1}$. The $\underline{1}$ corresponds to the singlet in Fig. 5(a). The $\underline{4}$ corresponds to adding an orbital part to this singlet, i.e., multiplied by x^i. Since there are two such singlets there are two $\underline{5}$'s. On the Fermi side the two $\underline{4}$'s that are required come directly from the $\underline{4}$'s of spin. The two singlets are obtained by contracting the single index with that of x^i. The $\underline{5}$ is a single index tensor. We can add any number of symmetrized indices to this and get higher representations of O(5) as shown in Fig. 5(b). These would all be allowed since they just correspond, in the O(4) notation, to tacking on extra factors of $x^ix^j\cdots$. The next series of representations starts with the two index antisymmetric tensors $A^{ij} \sim \underline{10}$ of O(5). This breaks up into a $\underline{6} + \underline{4}$ [= (3,1) + (1,3) + (2,2)]. In the Bose sector the $\underline{6}$ is already present and the $\underline{4}$ can be obtained by contracting one of the indices with an x^i. In the Fermi sector the $\underline{4}$ is already present. The $\underline{6}$ can be obtained by multiplying by x^i and antisymmetrizing. The fact that there are two $\underline{4}$'s in the Fermi sector, but only one $\underline{6}$ in the Bose sector does not cause any problem. There is only one $\underline{10}$ and it contains a particular linear combination of the $\underline{4}$'s. The supersymmetry charge that takes the Bose state into the Fermi state is written as $Q = \psi^i_{1+i2} P_i$. Noting that $\underline{6}$ has a U(1) charge (that acts on Dirac indices) equal to zero, we see that Q transforms it into the linear combination of the $\underline{4}$'s that has the quantum numbers of $\psi^i_1 + \psi^i_2$. This U(1) charge is preserved by J, so the two 4's are not mixed by J, they merely transform simultaneously under J. This series that starts with a $\underline{10}$ can also be extended by adding any number of symmetrized indices corresponding to higher orbital excitations. Thus the spectrum is as shown in Fig. 6. These ideas can be easily followed through for any N and the result is given in Fig. 7.

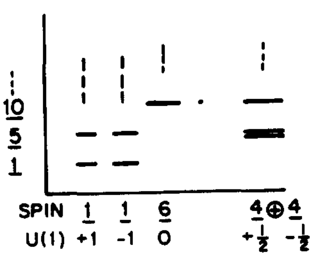


FIG. 6. The spectrum of the O(5) supersymmetric σ model in 0 + 1 dimension. The series are labeled by the dimension of the "spin" representation. The Dirac U(1) charge is also shown. The y axis has the dimension of the O(5) representation, which fixes the energy of the state.

We now turn to the $M = \frac{1}{2}$ supersymmetric sigma models. The construction that led to Table I can be repeated and we get the contents of Table II, the only difference being that the Dirac index is absent. One finds for odd N that both the Bose and Fermi states belong to the two spinor representations Σ_1 and Σ_2 of O($N-1$). The dimension of these representations is $2^{(N-3)/2}$ and accommodates all the states. Clearly the spinor representation of O(N) of dimension $2^{(N-1)/2}$ contains these two and satisfies all the other constraints. One can also add any number of symmetrized vector indices to get the higher orbital excitations. Further, since there are no singlets of O($N-1$), there is no zero energy state. The energy is once again given by the quadratic Casimir for which one can find expressions in the literature.[10,13] We do not reproduce them here since they are not particularly instructive. We have not worked out the case with N even but we expect by analogy to get the spinor representation.

A word on the analogy with the Wigner d functions. The key ingredient there was that $r \cdot J$, a position-dependent, radial, O(2) subgroup of O(3), had nonzero eigenvalues. Here it is S^{ab} or $\varepsilon^{abcd\cdots} n^a J^{bc}$, a "radial" O($N-1$) subgroup of O($N$) that has nonzero eigenvalues. Thus Wigner's d function is characterized by J^2 and $n \cdot J$, in our case it is J^2 and S^2. The analog of relation (9) there was $J^2 \geqslant (n \cdot J)^2$ (see Ref. 13). It must also be pointed out that in our case only specific (i.e., the antisymmetric or the spinor) representations of S

TABLE II. Transformation properties of the allowed $O(N)$ representations under the $O(N-1)$ spin subgroup for the case of the $M=\frac{1}{2}$ supersymmetry.

State	h_1	h_2	$\cdots$	h_i	$\cdots$	h_{N-1}	h_N	h_{α_1}	h_{α_2}	$\cdots$	$h_{\alpha_{N1}}$	$h_{\alpha N}$	$O(N-1)$	Rep.
$\|0\rangle$	$\frac{1}{2}$	$\frac{1}{2}$	$\cdots$	$\frac{1}{2}$	$\cdots$	$\frac{1}{2}\cdots$	$\frac{1}{2}$	0	0		0	1	Σ_1	$(=2^{(N-3)/2})$
$\psi^{J+}\|0\rangle$	$\frac{1}{2}$	$\frac{1}{2}$	$\cdots$	$\frac{1}{2}$	$\cdots$	$-\frac{1}{2}$	$\frac{1}{2}$	0	0		1	0	Σ_2	$(=2^{(N-3)/2})$
$\psi^{J+}\psi^{J+}\|0\rangle$														
$\vdots$				$\vdots$										

are allowed. A further generalization of the concept of Wigner's d function would presumably allow other representations of S also.

IV. CONCLUSION

In this paper we have computed the exact spectrum of the supersymmetric $O(N)$ σ models in $0+1$ dimension. It is tempting to conjecture that similar techniques can be used to solve the supersymmetric CP^{N-1} models[14] also.

ACKNOWLEDGMENT

This work was supported in part by the Department of Energy Contract No. EY-76-C-02-3071.

[1]B. Sathiapalan, Phys. Rev. Lett. **55**, 669 (1985).

[2]E. Witten, Phys. Rev. D **16**, 2991 (1977).

[3]J. Wess and B. Zumino, Phys. Lett. B **37**, 95 (1971); E. Witten, Nucl. Phys. B **223**, 422 (1983).

[4]A. C. Davis, A. J. MacFarlane, and J. W. van Holten, Nucl. Phys. B **216**, 394 (1983).

[5]A. C. Davis, A. J. MacFarlane, P. Popat, and J. W. van Holten, Cambridge preprint DAMTP 8417.

[6]P. A. M. Dirac, *Lectures on Quantum Mechanics* (Belfer Graduate School of Science, New York, 1964), Monograph No. 2.

[7]P. Casalbuoni, Nuovo Cimento A **33**, 389 (1976).

[8]E. Witten, Nucl. Phys. B **202**, 253 (1982); L. Alvarez-Gaume, Commun. Math. Phys. **90**, 161 (1983). See, for instance, T. Eguchi, P. B. Gilkey, and A. J. Hanson, Phys. Rep. **213**, 215 (1980).

[9]B. C. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1976).

[10]Another way to see this is to work with the combination ψ^i_{1+i2} and ψ^i_{1-i2}. In this basis it is easy to see that one gets the n-index antisymmetric representations of the $O(N-1)$ group. This method has the advantage that it works even when N is even. However, it does not work for the case of $M=\frac{1}{2}$ supersymmetry.

[11]One could try to see if relations analogous to (8) can be derived for the quartic and higher Casimir operators of $O(N)$. This does not seem to be the case, however, since mixed terms of the form LLSS that are not positive definite do not vanish.

[12]A. M. Perelmov and V. Popov, JETP Lett. **2**, 20 (1965).

[13]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, NJ, 1960).

[14]A. D'Adda, P. Di Vecchia, and M. Luscher, Nucl. Phys. B **152**, 125 (1979).

The Schwinger–DeWitt proper-time method for odd-dimensional gauge theory

Taejin Lee
Department of Physics, FM-15, University of Washington, Seattle, Washington 98195

(Received 5 March 1986; accepted for publication 7 May 1986)

The Schwinger–DeWitt proper-time method (WKB expansion) is applied to calculate the anomaly in odd-dimensional gauge theories. The parity violating part of effective action for gauge theory in odd dimensions with massless fermion is calculated explicitly and efficiently by this method. It is shown to be precisely the local Chern–Simons term.

I. INTRODUCTION

The Schwinger–DeWitt proper-time method[1] (WKB expansion) is one of the efficient and elegant techniques to study chiral anomalies[2] in even dimension, stress-tensor trace anomalies,[3] renormalization of the effective action[4] in flat or curved space-time. Meanwhile anomalies in odd dimensions have recently attracted much attention just as anomalies in even-dimensional gauge theories did previously.[5] We apply the Schwinger–DeWitt proper-time method to evaluate the anomaly in an odd-dimensional gauge theory. (By use of the original Schwinger proper-time method, a calculation for a homogeneous electromagnetic field in three dimensions has been carried out by Redlich.[6]) Anomalies in odd-dimensional gauge theories are intimately related to anomalies in even-dimensional gauge theories.

Let us compare two anomalies briefly. The chiral $U(1)_A$ anomaly in $2n$ dimensions is expressed by a topological invariant known as the nth Chern characteristic[7]

$$\Omega_{2n}[A] = \mathrm{gtr}\, F^n, \tag{1}$$

where F is a two-form given by[8]

$$F = dA + A^2 = \tfrac{1}{2} F_{\mu\nu} T^a\, dx^\mu\, dx^\nu \tag{2}$$

and gtr is the trace over group indices. The anomaly in an odd-dimensional gauge theory is the parity violating part of an induced effective action that spontaneously breaks down the symmetry of classical theory, that is, parity. In $(2n-1)$-dimensional gauge theory, it is given by a topological invariant, the so-called secondary characteristic class[9]

$$\omega_{2n-1}[A] = n \int_0^1 dt\, [A(t\, dA + t^2 A^2)^{n-1}]. \tag{3}$$

The intimate relationship between these two topological invariants is well-expressed mathematically as

$$\Omega_{2n}[A] = d\omega_{2n-1}[A]. \tag{4}$$

In fact, the fractional fermion number[10] induced through the parity violating part of low-energy effective action is obtained by use of Euclidean chiral anomaly equation in one lower dimension. This point was discussed more extensively for gauge anomalies and gravitational anomalies[11] by Alvarez-Gaumé *et al.*

One more interesting point is that the radiative corrections from higher-order loops to anomaly in $(2+1)$-dimensional Abelian gauge theory are absent as no corrections from higher-order loops to triangular anomaly in $(3+1)$-dimensional Abelian gauge theory arise.[12] With a vanishing tree level topological mass for an Abelian gauge field, no correction arises at two loops.[13] For the more general case, even if the Lagrangian has a nonvanishing bare topological mass term,[14] two loops do not contribute to the topological mass for the Abelian gauge field. It is also proved that the coefficient of topological mass term at one loop order is exact up to all higher orders[15] for $(2+1)$-dimensional Abelian gauge theory by use of the path integral formalism and by an analysis of Feynman diagrams.[16]

The parity anomaly in $(2+1)$-dimensional quantum electrodynamics is especially important, since it may be relevant to the fractionally quantized Hall effects.[17]

In the following sections, we evaluate the parity violating part of the effective action for the gauge field (the parity odd part of Fermion determinant for an arbitrary background gauge field) explicitly and efficiently by the Schwinger–DeWitt proper-time method.

II. THE SCHWINGER–DEWITT PROPER-TIME METHOD FOR ODD-DIMENSIONAL GAUGE THEORY

We begin by briefly discussing applications of the Schwinger–DeWitt proper-time method to even-dimensional gauge theories. This method produces a useful evaluation of the fermion determinant, which is the contribution to the effective action for the gauge field, induced by the fermion loops. Thus, it yields the simple calculation of the chiral anomaly in an even-dimensional gauge theory. The fermion determinant or the effective action is

$$\int L_{\mathrm{eff}} = -i \ln \mathrm{Det}(\not{D} + m). \tag{5}$$

We may compactify the Euclidean space-time by assuming that the gauge field vanishes sufficiently fast at infinity. (The anomaly in an odd-dimensional manifold with a boundary will be discussed in a separate note.[18]) Thanks to the existence of γ^5 in even dimensions, it can be cast in the following form:

$$\int L_{\mathrm{eff}} = -\frac{i}{2} \ln \mathrm{Det}(\not{D} + m)(\not{D} - m)$$
$$= -\frac{i}{2} \mathrm{Tr} \ln\Big(D^2 + \frac{e}{2}\sigma\cdot F - m^2\Big), \tag{6}$$

where Tr means trace over all indices including spinor indices, group indices, and space-time coordinates.

Rewriting it in the proper-time integral representation

$$\int L_{\text{eff}} = \frac{i}{2}\int_0^\infty \frac{d\tau}{\tau}\,\mathrm{Tr}\,e^{-\tau(H^2+m^2)}, \tag{7}$$

then we can define a Hamiltonain $H = -D^2 - (e/2)\sigma\cdot F$ and a Hilbert space. Now what we have to do is to evaluate a heat kernel, which is defined by $\langle x\tau|y\rangle = \langle x|e^{-\tau H}|y\rangle$. The heat kernel satisfies a Schrödinger equation

$$\frac{\partial}{\partial\tau}\langle x\tau|y\rangle = -H\langle x\tau|y\rangle \tag{8}$$

and has an adiabatic expansion (WKB expansion) in the coincidence limit:

$$\lim_{x\to y}\langle x|e^{-\tau H}|y\rangle = \frac{1}{(4\pi\tau)^{n+1/2}}\sum_{j=0}^{\infty} a_j(x)\tau^j. \tag{9}$$

Here n is dimension of space-time. The coefficients $a_j(x)$ can be obtained by the use of recursion relations that can be given by the substitution of Eq. (9) into Eq. (8).

Returning to odd-dimensional gauge theories, we immediately realize that this method cannot be applied straightforwardly to the evaluation of the fermion determinant in odd dimensions. In odd dimensions, there is no γ^5, hence the form of determinant cannot be cast in the convenient one (6). Instead, we decompose the effective action into the parity conserving part and the parity violating part:

$$\begin{aligned}\int L_{\text{eff}} &= -\frac{i}{2}\mathrm{Tr}[\ln(\not{D}+m)+\ln(\not{D}-m)]\\ &\quad -\frac{i}{2}\mathrm{Tr}[\ln(\not{D}+m)-\ln(\not{D}-m)]\\ &= \Gamma_c + \Gamma_p\end{aligned} \tag{10}$$

(in odd dimensions, the fermion mass term is parity odd). The Schwinger–DeWitt proper-time method can be applied straightforwardly to the parity even part. The parity odd part needs care. The strategy is to take the variation with respect to the gauge field, that is, the induced current from the parity odd part of the effective action:

$$\begin{aligned}\frac{\delta\Gamma_p}{\delta A_a^\mu} &= -\frac{i}{2}\mathrm{tr}\left[\frac{1}{\not{D}+m}-\frac{1}{\not{D}-m}\right]\gamma_\mu T^a\\ &= im\,\mathrm{tr}\left[\frac{1}{D^2+(e/2)\sigma\cdot F-m^2}\right]\gamma_\mu T^a,\end{aligned} \tag{11}$$

where tr is trace over spinor and group indices. If we rewrite also (11) in the proper-time integral representation,

$$\begin{aligned}\frac{\delta\Gamma_p}{\delta A_a^\mu} &= -im\,\mathrm{tr}\int_0^\infty d\tau\, e^{-\tau(H+m^2)}\gamma_\mu T^a,\\ H &= -D^2-(e/2)\sigma\cdot F.\end{aligned} \tag{12}$$

Thus once the heat kernel (9) is calculated, the evaluation of the induced current from the parity odd effective action is a simpler matter. The parity odd part of the effective action can be obtained by an integral of the induced current over gauge field.

In the following section, we will calculate the parity odd part of the effective action explicitly by the Schwinger–DeWitt proper-time method discussed in this section.

III. ANOMALY IN ODD DIMENSIONS

In some theories, the symmetries of classical theories are broken spontaneously in their quantum theories. In general, we will call these phenomena anomalies. In an odd-dimensional gauge theory, one of the anomalies is the spontaneous breakdown of the parity through the contribution of quantum fermion loops to the effective action for the gauge field.

The classical action for a massless fermion in odd dimensions is invariant under parity transformation. Under parity transformation,

$$\begin{aligned}&x\to\bar{x} = (x_0,-x_i),\quad (A_0,A_i)(x)\to(A_0,-A_i)(\bar{x}),\\ &\psi(x)\to i\gamma^0\psi(\bar{x}),\quad \int(\bar{\psi}\not{D}\psi)(x)\to\int(\bar{\psi}\not{D}\psi)(\bar{x}).\end{aligned} \tag{13}$$

Thus parity is a good symmetry at the classical level. But the quantum theory is not yet defined until the ultraviolet and infrared divergences have been regulated. These regulators may break the classical parity symmetry. Infrared divergence may be associated with the zero modes of the Dirac operator, if the fermion mass is vanishing, hence special care is needed for the infrared divergence problem in odd-dimensional gauge theories.[19] We will not discuss here the infrared divergence problem, which is the other side of the long story, and will concentrate only on the regularization for ultraviolet divergence. We will choose the Pauli–Villar regulator for ultraviolet divergence and examine its consequence, that is, anomaly.

The induced current from the parity odd part of the effective action regularized by the Pauli–Villar method is

$$\frac{\delta\Gamma_p}{\delta A_a^\mu} = -iM\,\mathrm{tr}\left[\frac{1}{D^2+(e/2)\sigma\cdot F-M^2}\right]\gamma_\mu T^a, \tag{14}$$

where M is an arbitrarily large mass for the regulator.

As we mentioned before, in the coincidence limit, the heat kernel has an adiabatic expansion (9) (WKB expansion). By making use of Eq. (9) and integrating over τ, we get a formal series expanded in $1/M$:

$$\begin{aligned}\frac{\delta\Gamma_p}{\delta A_a^\mu} &= \sum_j \frac{iM}{(M^2)^{j-n+1/2}}\frac{1}{(4\pi)^{n+1/2}}\\ &\quad\times\Gamma(j-n+\tfrac{1}{2})\,\mathrm{tr}\,a_j(x)\gamma_\mu T^a.\end{aligned} \tag{15}$$

Thus as $M\to\infty$, the coefficients of a_j $(n+1<j)$ do vanish. We just have the first finite number of terms. Now let us take a close look at $a_j(x)$ $(0<j<n)$, which may yield a divergent term in the effective action. The functions $a_j(x)$ $(0<j<n)$ may be obtained by using recursion relations that can be given by substitution of Eq. (9) into Eq. (8). But we will utilize a useful alternative form of the heat kernel by Nepomechi[20]:

$$\begin{aligned}\lim_{x\to y}\langle x|e^{-\tau H}|y\rangle &= \frac{1}{(4\pi\tau)^{n+1/2}}\int\frac{d^{2n+1}k}{\pi^{n+1/2}}e^{-k^2}\\ &\quad\times\sum_{j=0}^{\infty}\frac{1}{j!}[\tau^{1/2}2ik\cdot D-\tau H]^j.\end{aligned} \tag{16}$$

By comparing Eq. (16) and Eq. (9), we can read $a_j(x)$. Using a simple gamma algebra in $2n+1$ dimensions,

$$\mathrm{tr}\,\sigma_{\mu_1\nu_1}\cdots\sigma_{\mu_i\nu_i}\gamma_\mu = 0 \quad (\text{for } i\leqslant n-1),$$
$$\mathrm{tr}\,\sigma_{\mu_1\nu_1}\cdots\sigma_{\mu_n\nu_n}\gamma_\mu = i^{n+1}\epsilon_{\mu_1\nu_1\cdots\mu_n\nu_n\mu}, \tag{17}$$

we can see that $a_j(x)$ (for $j\leqslant n-1$) does not contribute to the right-hand side of Eq. (15) and only one term in $a_n(x)$ makes a nonvanishing contribution. We may get the precise evaluation of the induced current from the parity violating part, which is free of ultraviolet divergence,

$$\frac{\delta\gamma_p}{\delta A_\mu^a} = -\frac{1}{2}\left(\frac{i}{2\pi}\right)^n \frac{1}{n!}\frac{M}{|M|}\,\mathrm{gtr}\,F^n\,dx^\mu\,T^a, \tag{18}$$

where γ_p is a $(2n+1)$-form defined by $\int\gamma_p = \Gamma_p$. This is exactly the variation of Chern–Simons term in general odd dimensions with respect to gauge field

$$\frac{\delta\omega_{2n+1}}{\delta A_\mu^a} = (n+1)\,\mathrm{gtr}\,F^n\,dx^\mu\,T^a. \tag{19}$$

A simple integral over gauge field yields the parity violating part of the effective action, which is shown to be exactly the local Chern–Simons term up to a constant.

We will close this note with a comment on gauge theories with massive fermions. The variation of the parity odd part of the effective action with respect to the gauge field or the induced current has two terms

$$\frac{\delta\Gamma_p}{\delta A_a^\mu} = im\,\mathrm{tr}\left[\frac{1}{D^2+(e/2)\sigma\cdot F-m^2}\right]\gamma_\mu T^a - iM\,\mathrm{tr}\left[\frac{1}{D^2+(e/2)\sigma\cdot F-M^2}\right]\gamma_\mu T^a. \tag{20}$$

As we have shown, the first term of (20) also can be expanded formally in $1/m$ for a finite mass m. The leading term is the variation of a topological invariant, the nth Chern characteristic, with respect to the gauge field. Hence the low-energy effective action is (up to a constant)

$$\Gamma_p = \frac{1}{2}\frac{1}{(n+1)!}\left(\frac{i}{2\pi}\right)^n\left(\frac{m}{|m|}-\frac{M}{|M|}\right)\int\omega_{2n+1}[A]. \tag{21}$$

This agrees with the result in the Ref. 11 where the parity violating part of the effective action defined on a manifold $S^1\times\mathscr{M}^{2n}$ is evaluated by relating it with the even-dimensional gauge anomaly.

ACKNOWLEDGMENT

I would like to thank Professor D. G. Boulware for the helpful comment.

[1] J. Schwinger, Phys. Rev. **82**, 664 (1951); B. S. De Witt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965); Phys. Rep. **19**, 295 (1975).

[2] C. Lee, Nucl. Phys. B **202**, 336 (1982).

[3] L. S. Brown, Phys. Rev. D **15**, 1469 (1977); L. S. Brown and J. P. Cassidy, *ibid.* **15**, 2810 (1977); D. M. Capper and M. J. Duff, Nuovo Cimento A **23**, 173 (1973); Phys. Lett. A **53**, 361 (1975).

[4] N. D. Birrel and P. C. W. Davis, *Quantum Fields in Curved Space* (Cambridge U. P., Cambridge, 1982); B. S. DeWitt, in *Quantum Gravity II*, edited by C. Isham, R. Penrose, and D. Sciama (Oxford U. P., New York, 1981); D. G. Boulware, Phys. Rev. D **23**, 389 (1981); L. Parker, in *Recent Developments in Gravitation*, Carg'ese, 1978, edited by M. Levy and S. Deser (Plenum, New York, 1979).

[5] L. Alvarez-Gaumé and E. Witten, Nucl. Phys. B **234**, 269 (1984); L. Alvarez-Gaumé and P. Ginsparg, *ibid.* **243**, 449 (1984).

[6] A. N. Redlich, Phys. Rev. D **29**, 2366 (1984).

[7] B. Zumino, Y. S. Wu, and A. Zee, Nucl. Phys. B **239**, 427 (1984).

[8] A brief review for differential form can be found in the Appendix of Ref. 7.

[9] Y. S. Wu, Ann. Phys. (NY) **156**, 194 (1984).

[10] A. Niemi and G. Semenoff, Phys. Rev. Lett. **51**, 2077 (1983).

[11] L. Alvarez-Gaumé, S. Della Pietras, and G. Moore, Harvard University preprint HUTP-84/A028, 1984.

[12] S. Adler and W. A. Bardeen, Phys. Rev. **182**, 1517 (1969); see also A. Zee, Phys. Rev. Lett. **29**, 1198 (1972).

[13] Y. Kao and M. Suzuki, Phys. Rev. D **31**, 2167 (1985).

[14] M. Bernstein and T. J. Lee, Phys. Rev. D **32**, 1020 (1985).

[15] T. J. Lee, Phys. Lett. B **171**, 247 (1986).

[16] S. Coleman and B. Hill, Phys. Lett. B **159**, 184 (1985).

[17] R. Jackiw, Phys. Rev. D **29**, 2375 (1984); K. Ishikawa, Phys. Rev. Lett. **53**, 1615 (1984); Phys. Rev. D **31**, 1432 (1985).

[18] T. J. Lee (in preparation).

[19] R. Jackiw and S. Templeton, Phys. Rev. D **23**, 2291 (1981); T. Appelquist and R. D. Pisarski, *ibid.* **23**, 2305 (1981); M. de Roo and K. Stam, Nucl. Phys. B **246**, 335 (1984).

[20] R. Nepomechi, Phys. Rev. D **31**, 3291 (1985).

Mathematical aspects of quantum fluids. II. Nonrotating ^{4}He and Clebsch representations of symplectic two-cocycles

B. A. Kupershmidt
The University of Tennessee Space Institute, Tullahoma, Tennessee 37388 and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 30 July 1985; accepted for publication 30 April 1986)

For nonrotating and rotating ^{4}He, the formulas for the Poisson brackets and their canonical representations are shown to be particular cases of general Hamiltonian maps associated to symplectic two-cocycles on semidirect product Lie algebras of the type $\mathfrak{g} \ltimes (W \oplus V^* \oplus V)$.

I. INTRODUCTION

Superfluid helium has a very complex mathematical structure even on the macroscopic level of description. The ultimate reason for this complexity, at least for ^{4}He, can be traced to the fact that superfluid helium is a two-fluid system; in addition to that, for ^{3}He, new difficulties are caused by the presence of orbital angular momentum and various spins, all so(3) valued. The ways this complexity manifests itself vary greatly for four basic quantum fluids for which a tolerable mathematical description is known at the present time (see Refs. 1 and 2): nonrotating ^{4}He; rotating ^{4}He; spinless anisotropic ^{3}He–A; and anisotropic ^{3}He–A with spin. The goal of this series of papers of which the present one is the second, is *to interpret Lie algebraically* diverse mathematical facts observed in the description of these quantum fluids.

The subject I treat here is *Clebsch representations for systems whose Hamiltonian description involves a symplectic two–cocycle*. To be more precise, recall[2] the Poisson bracket formula for nonrotating ^{4}He:

$$\{H,F\} \sim \Big\{ \frac{\delta F}{\delta M_k}\Big[(M_l\partial_k + \partial_l M_k)\Big(\frac{\delta H}{\delta M_l}\Big) + \rho\partial_k\Big(\frac{\delta H}{\delta \rho}\Big) - \alpha_{,k}\frac{\delta H}{\delta\alpha} + \sigma\partial_k\Big(\frac{\delta H}{\delta\sigma}\Big)\Big] + \Big[\frac{\delta F}{\delta\rho}\partial_l\rho + \frac{\delta F}{\delta\alpha}\alpha_{,l} + \frac{\delta F}{\delta\sigma}\partial_l\sigma\Big]\Big(\frac{\delta H}{\delta M_l}\Big)\Big\} \tag{1.1a}$$

$$+\Big(\frac{\delta F}{\delta\alpha}\frac{\delta H}{\delta\rho} - \frac{\delta F}{\sigma\rho}\frac{\delta H}{\delta\alpha}\Big). \tag{1.1b}$$

The notation here is $\partial_l = \partial/\partial x_l$, where $(x_1,...,x_n)$ are coordinates in $\mathbf{R}^n$ ($n=3$ in physics); $(\cdot)_{,l} = \partial(\cdot)/\partial x_l$, $1 < k,l < n$, and the sum is taken over repeated indices; $\mathbf{M} = (M_1,...,M_n)$ is the total momentum density (of the normal flow); ρ is the mass density; σ is the entropy density; α is the phase of the order parameter which defines the curl-free superfluid velocity $\mathbf{v}^s$ as $\mathbf{v}^s = \nabla\alpha$; $\delta H/\delta(\cdot)$ denotes the variational derivative of H with respect to $(\cdot)$; and $\sim$ means equality modulo total derivatives (or "divergences"). The part (1.1a) of the Poisson bracket (1.1) is the natural bracket associated to the dual space of the semidirect product Lie algebra

$$\mathfrak{g}(^4\mathrm{He}_{nr}) = D_n \ltimes [\Lambda^0 \oplus \Lambda^n \oplus \Lambda^0], \tag{1.2}$$

with the commutator

$$[(X;f;\beta;a),(\bar X;\bar f;\bar\beta;\bar a)] = ([X,\bar X];X(\bar f) - \bar X(f);X(\bar\beta) - \bar X(\beta);X(\bar a) - \bar X(a)), \tag{1.3}$$

where D_n is the Lie algebra of vector fields on $\mathbf{R}^n$; $\Lambda^k = \Lambda^k(\mathbf{R}^n)$ is the $C^\infty(\mathbf{R}^n)$-module of differential k-forms on $\mathbf{R}^n$; $X,\bar X \in D^n$; $f,a,\bar f,\bar a \in \Lambda^0$; $\beta,\bar\beta\in\Lambda^n$; the (Lie derivative) action of D_n on Λ^k is denoted $X(\cdot)$ for $X\in D_n$ and $(\cdot)\in\Lambda^k$; and the dual coordinates on $(\mathfrak{g}(^4\mathrm{He}_{nr}))^*$ are M_k to $\partial_k \in D_n$, ρ to $1\in\Lambda^0$, α to $dx_1\wedge\cdots\wedge dx_n\in\Lambda^n$, and σ to $1\in\Lambda^0$. The part (1.1b) of the Poisson bracket (1.1) corresponds to the following two-cocycle on the Lie algebra $\mathfrak{g}(^4\mathrm{He}_{nr})$ (1.2):

$$\omega((X;f;\beta;a),(\bar X;\bar f;\bar\beta;\bar a)) = -f\bar\beta + \beta\bar f. \tag{1.4}$$

The nature of the two-cocycle (1.4) on the Lie algebra (1.2) has been explained in the first paper of this series (see Ref. 3, Theorem 3.1 and Remark 3.2): If $\mathfrak{g}$ is a Lie algebra over a function ring K, V_i, $i=1,2,3$, are vector spaces over K, θ: $V_1\times V_2 \to K$ is a bilinear differential operator, π_i: $\mathfrak{g}\to\mathrm{Diff}(V_i)$, $i=1,2$, are θ-adjoint representations of $\mathfrak{g}$, i.e.,

$$\theta(\pi_1(X)(v_1),v_2) + \theta(v_1,\pi_2(X)(v_2)) \sim 0,$$

$\forall X\in\mathfrak{g}$, $\forall v_i\in V_i$, then on the semidirect product Lie algebra $\mathfrak{g}\ltimes(V_1\oplus V_2\oplus V_3)$ the following expression is a (generalized) symplectic two-cocycle

$$\omega((X;v_1;v_2;v_3),(X';v_1';v_2';v_3')) = \theta(v_1',v_2) - \theta(v_1,v_2')).$$

(I shall be more precise in the main body of the paper.) Taking $\mathfrak{g} = D_n$, $V_1 = V_3 = \Lambda^0$, $V_2 = \Lambda^n$, and $\theta(v_1,v_2) = v_1v_2$, one recovers formulas (1.2)–(1.4).

Formula (1.1) was obtained in Ref. 2 by a direct mathematical computation of the following type: Consider a symplectic space Ψ with canonically conjugate pairs of variables $(\rho;\alpha)$, $(\sigma;\beta)$, $(f^k;\gamma^k)$, $k=1,...,m$. Then the map

$$M_i = \rho\alpha_{,i} + \sigma\beta_{,i} + \sum_u f^k\gamma^k_{,i}, \quad 1\leqslant i\leqslant n,$$

$$\rho=\rho,\quad \alpha=\alpha,\quad \sigma=\sigma, \tag{1.5}$$

is Hamiltonian between the symplectic structure of the space Ψ and the Hamiltonian structure (1.1). [If we were working in a finite-dimensional situation, and had the number

 0022-2488/86/092437-08$02.50

$m > (n-1)/2$, we would have been able to talk about the "reduction" of the symplectic structure on the "submanifold (1.5)."] Hamiltonian maps from a symplectic space into the dual of a Lie algebra, as well as their generalizations into the functional case, are called *Clebsch representations*. A general theory of such maps was developed in Ref. 4, Chap. VIII, §4. However, that theory turns out to be not general enough since the maps of the type (1.5), and the symplectic two-cocycles of the type (1.4), are not covered by it. It is, then, the *purpose of this paper to develop a suitable generalization of the theory of Clebsch representations in order to cover the case of nonrotating* ^{4}He.

The main result, Theorem 3.2 below, can be formulated especially simply in the finite-dimensional situation (i.e., when everything is considered over a field instead of over a function ring). Let $\mathfrak{g}$ be a Lie algebra and π: $\mathfrak{g}\to \mathrm{End}(V)$, $\bar{\pi}$: $\mathfrak{g}\to\mathrm{End}(W)$ be its representations. Then $\mathfrak{g}$ also acts, by the dual representations, on V^* and W^*. The usual Clebsch representation

$$R: V\oplus V^*\oplus W\oplus W^*\to(\mathfrak{g}\ltimes(V\oplus W))^*$$

is given by the formula [formula VIII (4.4) in Ref. 4]

$$\begin{aligned}&\langle R(v\oplus v^*\oplus w\oplus w^*), l\ltimes(v'\oplus w')\rangle\\ &\quad=\langle v^*,v'-\pi(l)(v)\rangle+\langle w^*,w'-\bar{\pi}(l)(w)\rangle.\end{aligned}\tag{1.6}$$

The natural Hamiltonian structure associated to the Lie algebra $\mathfrak{g}\ltimes(V\oplus V^*\oplus W)$, together with the symplectic two-cocycle on the $V\oplus V^*$ part of it, can be obtained via the following more general Clebsch map $\bar{R}$, from the same symplectic space $V\oplus V^*\oplus W\oplus W^*$:

$$\begin{aligned}&\bar{R}^*(l\ltimes(v'\oplus v^{*\prime}\oplus w'))(v\oplus v^*\oplus w\oplus w^*)\\ &\quad=\langle v^*,v'-\pi(l)(v)\rangle+\langle w^*,w'-\bar{\pi}(l)(w)\rangle\\ &\qquad-\langle v^{*\prime},v\rangle,\end{aligned}\tag{1.7}$$

where $l\ltimes(v'\oplus v^{*\prime}\oplus w')$ is considered as a typical linear function on the space $(\mathfrak{g}\ltimes(V\oplus V^*\oplus W))^*$. Moreover, the injection

$$\Pi: l\ltimes(v'\oplus w')\mapsto l\ltimes(v'\oplus 0\oplus w')$$

can be easily seen to be a Hamiltonian map. Therefore, one immediately recovers the familiar Clebsch map R (1.6) from the more general formula (1.7) via the relation $R^*=\bar{R}^*\Pi$. This will be made precise in Sec III.

The plan of the presentation is as follows. In the next section, I summarize basic facts of the Hamiltonian formalism, including transformation formula (2.9) for the variational derivatives, criterium (2.16) for a map to be Hamiltonian, and one-to-one correspondence between affine Hamiltonian operators and generalized two-cocycles on differential–difference Lie algebras. In Sec. III, the main result of this paper (Theorem 3.2) is proved, providing a Clebsch representation for the symplectic two-cocycle on a semidirect product Lie algebra of the type $\mathfrak{g}\ltimes(W\oplus V^*\oplus V)$. In the last section (Sec. IV), this Theorem is applied to two types of ^{4}He: the nonrotating ^{4}He, described by formulas (1.1) and (1.5); and the rotating ^{4}He, described by formulas (4.6), (4.9), and (4.10).

The reader accustomed to associating two-cocycles on a Lie algebra with central extensions of this Lie algebra should take care to note that the two-cocycles appearing in the description of quantum fluids and with which we work in this paper [formula (2.23) below] are *generalized* ones and have nothing to do with central extensions or representation theory.

II. HAMILTONIAN STRUCTURES AND HAMILTONIAN MAPS

In this section the basic formulas from the calculus of variations and the Hamiltonian formalism are recalled. Details may be found in Refs. 3 and 4.

Let K be a commutative algebra. Let $\partial_1,\dots,\partial_n: K\to K$ be n commuting derivations. Let G be a discrete group acting by automorphisms on K, and suppose that the actions of G and ∂'s commute. Such a K is called a differential–difference ring. (In the absence of G, K can be thought of as the ring of smooth functions on $\mathbf{R}^n$; in the absence of ∂'s, K can be thought of as the ring of functions on G.) Let I be a countable set. Set $C=K[q_i^{(g|\nu)}]$, $i\in I$, $g\in G$, $\nu\in\mathbf{Z}^n_+$ (polynomials in the variables $q_i^{(g|\nu)}$ with coefficients in K), and extend G and the ∂'s to act on C by the rule

$$\hat{h}(q_i^{(g|\nu)})=q_i^{(hg|\nu)},\quad \partial^\mu(q_i^{(g|\nu)})=q_i^{(g|\mu+\nu)},\qquad h\in G,\quad \mu\in\mathbf{Z}^n_+,\tag{2.1}$$

where $(\pm\partial)^\mu=(\pm\partial_1)^{\mu_1}\cdots(\pm\partial_n)^{\mu_n}$ for $\mu=(\mu_1,\dots,\mu_n)\in\mathbf{Z}^n_+$, and $\hat{h}(\cdot)$ is the image of $(\cdot)$ under the automorphism $\hat{h}$, $h\in G$.

Let N be a natural number or ∞, T a differential–difference ring. T^N consists of column vectors with only finite number of nonzero components. An operator E: $T^N\to T^N$ is a map of the form

$$(E(u))_j=\sum E^j_{i,g,\mu}\hat{g}\,\partial^\mu(u_i),\quad E^j_{\dots}\in T,\quad u\in T^N,\tag{2.2}$$

finite sums; a bilinear operator $T^{N_1}\times T^{N_2}\to T^{N_3}$, is defined analogously. An algebra structure on T^N is a bilinear operator $T^N\times T^N\to T^N$. The associative ring of operators $T^N\to T^N$, and the corresponding Lie algebra, are both denoted $\mathrm{Diff}(T^N)$. Trivial elements in T are defined as elements from

$$\mathrm{Im}\,\mathscr{D}=\sum_{g\in G}\mathrm{Im}(\hat{g}-\hat{e})+\sum_{s=1}^n\mathrm{Im}\,\partial_s,$$

where e is the unit element of G; we write $a\sim b$ if $(a-b)$ is trivial.

A derivation X of C over K is called evolutionary if it commutes with the actions of G and ∂'s, so that

$$X=\sum\hat{g}\partial^\nu(X_i)\frac{\partial}{\partial q_i^{(g|\nu)}},\quad X_i:=X(q_i),\quad q_i:=q_i^{(e|0)}.\tag{2.3}$$

The set of all evolution derivations is a Lie algebra denoted $D^{\mathrm{ev}}(C)$.

Set $N=|I|$. The Euler–Lagrange map $\delta=\delta/\delta\bar{q}$: $C\to C^N$, defined by the formula

$$\left(\frac{\delta H}{\delta\bar{q}}\right)_i=\frac{\delta H}{\delta q_i}=\sum\hat{g}^{-1}(-\partial)^\nu\left(\frac{\delta H}{\delta q_i^{(g|\nu)}}\right),\tag{2.4}$$

annihilates Im $\mathscr{D}$ in C:

$$\operatorname{Ker}\delta = \operatorname{Im}\mathscr{D} + K. \tag{2.5}$$

Here $\delta H/\delta q_i$ is called the variational derivative of H with respect to q_i. For $X\in D^{ev}(G)$, $H\in C$,

$$X(H) \sim \overline{X}^t \frac{\delta H}{\delta \overline{q}} \quad \text{(formula for the first variation)}, \tag{2.6}$$

where t stands for transpose, and

$$(\overline{X})_i = X_i. \tag{2.7}$$

The variational map δ has the following transformation properties. Let $C_1 = K[p_j^{(g|\nu)}]$, $j\in J, g\in G, \nu\in Z_+^n$, be another differential–difference ring. A (differential–difference) homomorphism Φ: $C\to C_1$ is a homomorphism over K, which commutes with the actions of G and the ∂'s:

$$\Phi(q_i^{(g|\nu)}) = \hat{g}\partial^\nu(\Phi_i), \quad \Phi_i := \Phi(q_i). \tag{2.8}$$

Let $H\in C$. The variational derivatives of H and $\Phi(H)$ are related by the formula

$$\frac{\delta\Phi(H)}{\delta\overline{p}} = D(\overline{\Phi})^\dagger \Phi\left(\frac{\delta H}{\delta\overline{q}}\right), \tag{2.9}$$

where

$$\overline{\Phi}\in C_1^N, \quad (\overline{\Phi})_i = \Phi(q_i), \tag{2.10}$$

$D(\overline{\Phi})$ is the Fréchet derivative of $\overline{\Phi}$,

$$D(\overline{\Phi})_{ij} = D_{p_j}(\Phi_i) = \sum \frac{\partial\Phi_i}{\partial p_j^{(g|\nu)}}\hat{g}\,\partial^\nu, \tag{2.11}$$

and $\dagger$ stands for adjoint:

$$(D(\overline{\Phi})^\dagger)_{ij} = [D(\overline{\Phi})_{ji}]^\dagger = \sum \hat{g}^{-1}(-\partial)^\nu \frac{\partial\Phi_j}{\partial p_i^{(g|\nu)}}. \tag{2.12}$$

We now move on to the Hamiltonian formalism. A map Γ: $C\to D^{ev}(C)$, $H\mapsto X_H$, is called Hamiltonian if there exists an operator B: $C^N\to C^N$ such that

$$\overline{X}_H = B\left(\frac{\delta H}{\delta\overline{q}}\right), \tag{2.13}$$

$$\{H,F\} \sim -\{F,H\} \quad \text{(skew symmetry)}, \tag{2.14}$$

where the Poisson bracket $\{H,F\}$ is defined as $X_H(F)$;

$$X_{\{H,F\}} = [X_H, X_F], \tag{2.15}$$

or equivalently,

$$\{H,\{F,S\}\} + \text{c.p.} \sim 0, \tag{2.15$'$}$$

for any $H,F,S\in C' = K'[q_i^{(g|\nu)}]$, with arbitrary (differential–difference) extension $K'\supset K$; here c.p. stands for cyclic permutation. The property (2.14) is equivalent to B being skew symmetric ($B^\dagger = -B$), while (2.15) can be reduced to a set of quadratic equations on the matrix elements of B.

If Γ_1: $C_1\to D^{ev}(C_1)$, $F\mapsto X_F$ is a Hamiltonian structure in the ring C_1, then the map Φ is called Hamiltonian (also "canonical") if, for any $H\in C$, the evolution derivations X_H in C and $X_{\Phi(H)}$ in C_1 are Φ compatible: $\Phi X_H = X_{\Phi(H)}\Phi$. If B_1 is such that $\overline{X}_F = X_F(\overline{p}) = B_1(\delta F/\delta\overline{p})$, then, using (2.9), one can show that Φ is Hamiltonian if and only if

$$\Phi(B) = D(\overline{\Phi})B_1D(\overline{\Phi})^\dagger. \tag{2.16}$$

A Lie algebra structure on K^N is an operator $K^N\times K^N\to K^N$, [,]: $X\times Y\mapsto[X,Y]$, satisfying the following conditions:

(i) $[X,Y] = -[X,Y]$ (skew symmetry),

(ii) $[X,[Y,Z]] + \text{c.p.} = 0$ (Jacobi identity), (2.17)

(iii) the properties (2.17) (i) and (ii) remain true under any (differential–difference) extension $K'\supset K$.

A bilinear form on K^N is an operator $K^N\times K^N\to K$. To each bilinear form ω one uniquely associates an operator b_ω: $K^N\to K^N$ acting by the rule

$$\omega(X,Y) \sim X^t b_\omega(Y), \tag{2.18}$$

so that if

$$\omega(X,Y) = \sum \omega_{i,g,\mu|j,h,\nu}\hat{g}\,\partial^\mu(X_i)\cdot\hat{h}\,\partial^\nu(Y_j), \quad \omega_{\dots}\in K,$$

then ("integrating by parts")

$$(b_\omega)_{ij} = \sum \hat{g}^{-1}(-\partial)^\mu \omega_{i,g,\mu|j,h,\nu}\hat{h}\,\partial^\nu. \tag{2.19}$$

The form ω is called symmetric (resp. skew symmetric) if $\omega(X,Y)\sim\omega(Y,X)$ [resp. $\omega(X,Y)\sim -\omega(X,Y)$]. The form ω is symmetric (resp. skew symmetric) if and only if the corresponding operator b_ω is symmetric, $(b_\omega)^\dagger = b_\omega$ [resp. b_ω is skew symmetric, $(b_\omega)^\dagger = -b_\omega$]. Recall that for an operator E: $K^N\to K^M$, the adjoint operator $E^\dagger$: $K^M\to K^N$ is uniquely defined by the equation

$$v^tE(u) \sim [E^\dagger(v)]^t_u, \quad \forall u\in K^N, \quad \forall v\in K^M, \tag{2.20}$$

so that

$$(E^\dagger)_{ij} = (E_{ji})^\dagger \tag{2.21}$$

and

$$(a\hat{g}\,\partial^\nu)^\dagger = (-\partial)^\nu\hat{g}^{-1}a, \quad a\in K. \tag{2.22}$$

A skew-symmetric form ω on a Lie algebra $\mathfrak{g} = K^N$ is called a (generalized) two-cocycle on $\mathfrak{g}$ if

$$\omega(X,[Y,Z]) + \text{c.p.} \sim 0, \quad \forall X,Y,Z\in\mathfrak{g}. \tag{2.23}$$

An operator $a\hat{g}\,\partial^\nu$: $C\to C$ is called q independent [resp. linear (in q)] if $a\in K$ (resp. if $a = \Sigma a_{i,h,\mu}q_i^{(h|\mu)}$, $a_{\dots}\in K$). An operator is affine if it is a sum of a q-independent and a q-linear operator. The same terminology applies to sums of operators, and to matrix operators.

Let $B = B^l + b$ be an affine operator: $C^N\to C^N$, with B^l being q linear and b being q independent. We make K^N into a (differential–difference) algebra setting

$$\overline{q}^t[X,Y] \sim X^tB^l(Y), \quad (\overline{q})_i = q_i, \quad X,Y\in K^N. \tag{2.24}$$

Conversely, given an algebra structure on K^N, (2.24) defines a q-linear operator B^l. The relation between affine Hamiltonian operators and two-cocycles on Lie algebras is one-to-one: given a Lie algebra $\mathfrak{g}$, and a two-cocycle ω on it, we set $B = B^l + b_\omega$, with B^l defined by (2.24). Conversely, given an affine Hamiltonian matrix $B = B^l + b$, the same formula (2.24) defines a Lie algebra structure on K^N while (2.18) defines a two-cocycle ω via $b_\omega = b$.

III. CLEBSCH REPRESENTATIONS FOR SYMPLECTIC TWO-COCYCLES

Let $\mathfrak{g} = K^N$ be a Lie algebra; $W = K^{M_1}$, $V = K^{M_2}$; ${}^1\rho$: $\mathfrak{g} \to \mathrm{Diff}(W)$, ${}^2\rho$: $\mathfrak{g} \to \mathrm{Diff}(V)$ be two representations of $\mathfrak{g}$. Representation ${}^3\rho(X) = -\rho^2(X)^\dagger, X \in \mathfrak{g}$, of $\mathfrak{g}$ on $V^*: = K^{M_2}$ is called the dual representation of $\mathfrak{g}$ on V^* (see Ref. 3, Proposition 3.3). Let $\bar{\mathfrak{g}} = \mathfrak{g} \ltimes (W \oplus V^* \oplus V)$ be the semidirect product of $\mathfrak{g}$ with $W \oplus V^* \oplus V$, and let B^l be the natural Hamiltonial matrix associated by (2.24) with the Lie algebra $\bar{\mathfrak{g}}$,

$$\begin{pmatrix} X \\ u_1 \\ u_3 \\ u_2 \end{pmatrix}^t B^l \begin{pmatrix} Y \\ v_1 \\ v_3 \\ v_2 \end{pmatrix}$$

$$\sim \sum q_k [X,Y]_k + \sum c_i (X \cdot v_1 - Y \cdot u_1)_i$$
$$+ \sum \gamma_j (X \cdot v_3 - Y \cdot u_3)_j$$
$$+ \sum \lambda_j (X \cdot v_2 - Y \cdot u_2)_j, \tag{3.1}$$

in the ring

$$C = K\,[q_k^{(g|\nu)}, c_i^{(g|\nu)}, \gamma_j^{(g|\nu)}, \lambda_j^{(g|\nu)}],$$
$$1 \leqslant k \leqslant N, \quad 1 \leqslant i \leqslant M_1, \quad 1 \leqslant j \leqslant M_2, \tag{3.2}$$

which plays the role of the functions on the "dual space to $\bar{\mathfrak{g}}$," where

$$X,Y \in \mathfrak{g}, \quad u_1, v_1 \in W, \quad u_2, v_2 \in V, \quad u_3, v_3 \in V^*,$$

and $X \cdot (\ )_i := {}^i\rho(X)(\ )$. We let ω be the symplectic two-cocycle on $\bar{\mathfrak{g}}$, with

$$b_\omega = \left(\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{array}\right), \tag{3.3}$$

and let the corresponding affine Hamiltonian structure B in C be defined as

$$B = B^l + b_\omega. \tag{3.4}$$

The role of "functions on $W \oplus W^* \oplus V \oplus V^*$" is played by the ring

$$C_1 = K\,[a_i^{(g|\nu)}, b_i^{(g|\nu)}, \varphi_j^{(g|\nu)}, \psi_j^{(g|\nu)}],$$
$$1 \leqslant i \leqslant M_1, \quad 1 \leqslant j \leqslant M_2; \tag{3.5}$$

the symplectic Hamiltonian matrix in C_1 is

$$B_1 = \left(\begin{array}{cc|cc} 0 & 1 & & \\ -1 & 0 & \multicolumn{2}{c}{0} \\ \hline & & 0 & 1 \\ \multicolumn{2}{c|}{0} & -1 & 0 \end{array}\right). \tag{3.6}$$

To describe the desired Clebsch map Φ: $C \to C_1$, we need some preparation.

Lemma 3.1: Let ρ: $\mathfrak{g} \to \mathrm{Diff}(T)$ be a representation, with

$$\rho(X)_{\alpha\beta} = \sum \rho_{\alpha\beta}^{k;g,\nu|h,\sigma} X_k^{(h|\sigma)} \hat{g} \partial^\nu, \quad \rho_{\dots}^{k\dots} \in K. \tag{3.7}$$

Define the map ∇: $T \times T \to \mathfrak{g}$, $u \times v \mapsto u \nabla v$, by

$$(u \nabla v)_k = \sum \hat{h}^{-1} (-\partial)^\sigma (\rho_{\alpha\beta}^{k;g,\nu|h,\sigma} v_\alpha u_\beta^{(g|\nu)}). \tag{3.8}$$

Then

$$v^t \rho(X)(u) \sim X^t (u \nabla v), \quad u,v \in T, \quad X \in \mathfrak{g}. \tag{3.9}$$

Proof: We have, by (3.7),

$$v^t \rho(X)(u) = \sum v_\alpha \, \rho_{\alpha\beta}^{k;g,\nu|h,\sigma} X_k^{(h|\sigma)} u_\beta^{(g|\nu)}$$

$$\sim \sum X_k \hat{h}^{-1} (-\partial)^\sigma (\rho_{\alpha\beta}^{k;g,\nu|h,\sigma} v_\alpha u_\beta^{(g|\nu)}). \quad \blacksquare$$

We use this lemma for the cases $\rho = {}^1\rho$, $T = W$ and $\rho = {}^2\rho$, $T = V$ to define

$$\Phi_k^1 = (a \nabla b)_k, \tag{3.10}$$
$$\Phi_k^2 = (\varphi \nabla \psi)_k. \tag{3.11}$$

Theorem 3.2: Define the following homomorphism Φ: $C \to C_1$:

$$\begin{gathered} \Phi(q_k) = \Phi_k^1 + \Phi_k^2, \quad \Phi(c_i) = b_i, \\ \Phi(\gamma_j) = \varphi_j, \quad \Phi(\lambda_j) = \psi_j. \end{gathered} \tag{3.12}$$

Then the map Φ is Hamiltonian with respect to the Hamiltonian structures (3.4) in C and (3.6) in C_1.

Remark: The cocycleless Clebsch representation given in Ref. 4 can be obtained as a special case of Theorem 3.2, when V, V^*, γ, λ, φ, and ψ are all absent.

Proof: We have to check out the equality (2.16). Since, by (3.12),

$$D(\bar{\Phi}) = \begin{array}{c} \\ \Phi(q_k) \\ \Phi(c_i) \\ \Phi(\gamma_j) \\ \Phi(\lambda_j) \end{array} \left(\begin{array}{cccc} a_s & b_s & \varphi_r & \psi_r \\ \dfrac{D\Phi_k^1}{Da_s} & \dfrac{D\Phi_k^1}{Db_s} & \dfrac{D\Phi_k^2}{D\varphi_r} & \dfrac{D\Phi_k^2}{D\psi_r} \\ 0 & \delta_i^s & 0 & 0 \\ 0 & 0 & \delta_j^r & 0 \\ 0 & 0 & 0 & \delta_j^r \end{array}\right), \tag{3.13}$$

for the right-hand side of (2.16) we obtain

$$D(\overline{\Phi})B_1D(\overline{\Phi})^\dagger = \begin{array}{rc} & \begin{array}{cccc} \Phi(q_l) & \Phi(c_r) & \Phi(\gamma_s) & \Phi(\lambda_s) \end{array} \\ \begin{array}{c} \Phi(q_k) \\[3ex] \Phi(c_i) \\[3ex] \Phi(\gamma_j) \\[3ex] \Phi(\lambda_j) \end{array} & \begin{pmatrix} \eta_{kl} & \dfrac{D\Phi^1_k}{Da_r} & -\dfrac{D\Phi^2_k}{D\psi_s} & \dfrac{D\Phi^2_k}{D\varphi_s} \\[2ex] -\left(\dfrac{D\Phi^1_l}{Da_i}\right)^\dagger & 0 & 0 & 0 \\[2ex] \left(\dfrac{D\Phi^2_l}{D\psi_j}\right)^\dagger & 0 & 0 & \delta^s_j \\[2ex] -\left(\dfrac{D\Phi^2_l}{D\varphi_j}\right)^\dagger & 0 & -\delta^s_j & 0 \end{pmatrix} \end{array}, \tag{3.14}$$

where

$$\eta_{kl} = \sum\left[\frac{D\Phi^1_k}{Da_s}\left(\frac{D\Phi^1_l}{Db_s}\right)^\dagger - \frac{D\Phi^1_k}{Db_s}\left(\frac{D\Phi^2_l}{Da_s}\right)^\dagger\right] + \sum\left[\frac{D\Phi^2_k}{D\varphi_r}\left(\frac{D\Phi^2_l}{D\psi_r}\right)^\dagger - \frac{D\Phi^2_k}{D\psi_r}\left(\frac{D\Phi^2_l}{D\varphi_r}\right)^\dagger\right]. \tag{3.15}$$

We see at once that the lower-right corner of the matrix (3.14) represents its q-independent part, and it equals exactly the matrix $b_\omega = \Phi(b_\omega)$ in (3.3). We, thus, can disregard this constant block, and compare the rest of the matrix (3.14), which we denote by $\widetilde{B}$, to the matrix $\Phi(B^l)$. To show that $\widetilde{B} = \Phi(B^l)$, it is enough to show that the following relation is satisfied:

$$\begin{pmatrix} X \\ u_1 \\ u_3 \\ u_2 \end{pmatrix}^t \widetilde{B} \begin{pmatrix} Y \\ v_1 \\ v_3 \\ v_2 \end{pmatrix} \sim \begin{pmatrix} X \\ u_1 \\ u_3 \\ u_2 \end{pmatrix} \Phi(B^l) \begin{pmatrix} Y \\ v_1 \\ v_3 \\ v_2 \end{pmatrix}, \quad \forall X,Y \in \mathfrak{g}, \quad \forall u_1,v_1 \in W, \quad \forall u_2,v_2 \in V, \quad \forall u_3,v_3 \in V^*. \tag{3.16}$$

Since both $\widetilde{B}$ and $\Phi(B^l)$ are skew symmetric, we can rewrite (3.16), with the help of (3.1) and (3.12), in the form

$$\sum X_k \eta_{kl}(Y_l) \sim \sum (\Phi^1_k + \Phi^2_k)[X,Y]_k, \tag{3.17}$$

$$\sum u_{1i}\left(\frac{D\Phi^1_k}{Da_i}\right)^\dagger (Y_k) \sim \sum b_i (Y \cdot u_1)_i, \tag{3.18}$$

$$\sum u_{3j}\left(\frac{D\Phi^2_k}{D\psi_j}\right)^\dagger (Y_k) \sim -\sum \varphi_j (Y \cdot u_3)_j, \tag{3.19}$$

$$\sum u_{2j}\left(\frac{D\Phi^2_k}{D\varphi_j}\right)^\dagger (Y_k) \sim \sum \psi_j (Y \cdot u_2)_j, \tag{3.20}$$

and we can discard (3.20) since it has the same structure as (3.18).

To prove (3.17)–(3.19), we use the following formulas:

$$\frac{D\Phi^1_k}{Da_i} = \sum \hat{h}^{-1}(-\partial)^\sigma\, {}^1\rho^{k;g,\nu|h,\sigma}_{is} b_s \hat{g}\,\partial^\nu, \tag{3.21}$$

$$\frac{D\Phi^1_k}{Db_i} = \sum \hat{h}^{-1}(-\partial)^\sigma\, {}^1\rho^{k;g,\nu|h,\sigma}_{is} a^{(g|\nu)}_s, \tag{3.22}$$

$$(X \cdot a)_i = \sum \left(\frac{D\Phi^1_k}{Db_i}\right)^\dagger (X_k), \quad a \in W, \tag{3.23}$$

$$b^t(X \cdot v) \sim \sum v_i \left(\frac{D\Phi^1_k}{Da_i}\right)^\dagger (X_k), \quad v \in W, \tag{3.24}$$

which can be proved as follows. Formulas (3.21) and (3.22) are direct consequences of (3.8) and (3.10). Now, from (3.7),

$$(X \cdot a)_i = \sum {}^1\rho^{k;g,\nu|h,\sigma}_{is} X^{(h|\sigma)}_k a^{(g|\nu)}_s = \sum {}^1\rho^{k;g,\nu|h,\sigma}_{is} a^{(g|\nu)}_s \hat{h}\partial^\sigma (X_k)$$

$$= \sum [\hat{h}^{-1}(-\partial)^\sigma\, {}^1\rho^{k;g,\nu|h,\sigma}_{is} a^{(g|\nu)}_s]^\dagger (X_k)$$

$$= \sum \left(\frac{D\Phi_k^1}{Db_i}\right)^\dagger (X_k) \quad [\text{by } (3.22)],$$

which is (3.23). Finally, by (3.7),

$$b'(X\cdot v) = \sum b_s\, {}^1\rho_{is}^{k;g,v|h,\sigma} X_k^{(h|\sigma)} v_i^{(g|v)} \sim \sum v_i \hat{g}^{-1}(-\partial)^v\, {}^1\rho_{si}^{k;g,v|h,\sigma} b_s \hat{h}\partial^\sigma (X_k)$$

$$= \sum v_i \left(\frac{D\Phi_k^1}{Da_i}\right)^\dagger (X_k) \quad [\text{by } (3.21)],$$

which is (3.24).

Now, (3.18) is, up to notation, (3.24). On the other hand, rewriting (3.23) in the form

$$(Y\cdot\varphi)_j = \sum \left(\frac{D\Phi_k^2}{D\psi_j}\right)^\dagger (Y_k),$$

and multiplying by u_{3j} from the left, we obtain

$$\sum u_{3j}\left(\frac{D\Phi_k^2}{D\psi_j}\right)^\dagger (Y_k) = u_3^t (Y\cdot\varphi) \sim -\varphi^t (Y\cdot u_3)\,[\text{since } {}^3\rho(Y) = -{}^2\rho(Y)^\dagger]$$

and this is (3.19).

It remains to prove (3.17), which is equivalent to

$$\sum X_k \left[\frac{D\Phi_k^1}{Da_s}\left(\frac{D\Phi_l^1}{Db_s}\right)^\dagger - \frac{D\Phi_k^1}{Db_s}\left(\frac{D\Phi_l^1}{Da_s}\right)^\dagger\right](Y_l) \sim \sum \Phi_k^1 [X,Y]_k, \tag{3.25}$$

which can be seen as follows:

$$\sum \Phi_k^1 [X,Y]_k = \sum (a\nabla b)_k [X,Y]_k \quad [\text{by } (3.10)]$$

$$\sim b'([X,Y]\cdot a) \quad [\text{by } (3.9)]$$

$$= b'[X\cdot(Y\cdot a) - Y\cdot(X\cdot a)] \quad [\text{since } {}^1\rho \text{ is a representation}]$$

$$\sim \sum (Y\cdot a)_i \left(\frac{D\Phi_k^1}{Da_i}\right)^\dagger (X_k) - \sum (X\cdot a)_i \left(\frac{D\Phi_l^1}{Da_i}\right)^\dagger (Y_l) \quad [\text{by } (3.24)]$$

$$= \sum \left(\frac{D\Phi_l^1}{Db_i}\right)^\dagger (Y_l)\left(\frac{D\Phi_k^1}{Da_i}\right)^\dagger (X_k) - \sum \left(\frac{D\Phi_k^1}{Db_i}\right)^\dagger (X_k)\left(\frac{D\Phi_l^1}{Da_i}\right)^\dagger (Y_l) \quad [\text{by } (3.23)]$$

$$\sim \sum X_k \frac{D\Phi_k^1}{Da_i}\left(\frac{D\Phi_l^1}{Db_i}\right)^\dagger (Y_l) - \sum X_k \frac{D\Phi_k^1}{Db_i}\left(\frac{D\Phi_l^1}{Da_i}\right)(Y_l) \quad [\text{by } (2.20)], \qquad \blacksquare$$

and this is exactly the left-hand side of (3.25).

Corollary 3.3: In the absence of W, we obtain a Hamiltonian map

$$\Phi: K[q_k^{(g|v)}, \gamma_j^{(g|v)}, \lambda_j^{(g|v)}] \to K[\gamma_j^{(g|v)}, \lambda_j^{(g|v)}],$$

given by the formula $\Phi(q_k) = (\gamma\nabla\lambda)_k$, relating the symplectic space $V^*\oplus V$ with the symplectic two-cocycle on $\mathfrak{g}\ltimes(V^*\oplus V)$.

It would be interesting to find an analog of Theorem 3.2 for semidirect products of the type $\mathfrak{g}\ltimes(W\oplus V_1\oplus V_2)$, when $V_1 \neq V_2^*$, and with *generalized* symplectic two-cocycle

$$\omega((v_1;v_2);(v_1';v_2')) = \theta(v_1',v_2) - \theta(v_1,v_2')$$

for θ-adjoint representations of $\mathfrak{g}$ on V_1 and V_2. An example of this sort is $G=\{e\}$, $n=1$, $\mathfrak{g}=D_1$, $W=\{0\}$, $V_1=V_2=K$, $\theta=\partial$, so that $b_\omega = \left(\begin{smallmatrix}0 & \partial\\ \partial & 0\end{smallmatrix}\right)$, and ${}^{V_1}\rho(X) = {}^{V_2}\rho(X) = X\partial$. Then the map

$$\Phi: K[q^{(m)},u^{(m)},v^{(m)}] \to K[u^{(m)},v^{(m)}],$$

$$\Phi(q) = uv, \quad \Phi(u) = u, \quad \Phi(v) = v, \tag{3.26}$$

is Hamiltonian between the Hamiltonian structures in C and C_1 with the corresponding Hamiltonian matrices

$$B = \begin{pmatrix} q\partial + \partial q & u\partial & v\partial \\ \partial u & 0 & 0 \\ \partial v & 0 & 0 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 0 & \partial \\ \partial & 0\end{pmatrix}. \tag{3.27}$$

This can be seen at once by verifying formula (2.16).

On the other hand, the example below shows that for arbitrary θ-adjoint representations a Hamiltonian map should not be, in general, expected to exist.

Example: Again, $G=\{e\}$, $n=1$, $\mathfrak{g}=D_1$, $V_1=K$, $V_2=K^2$, with ${}^{V_1}\rho(X) = X\partial$,

$${}^{V_2}\rho(X) = X\partial\mathbf{1} + \partial(X)\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix},$$

$$\theta(v_1;(v_{21};v_{22})) = v_1[\partial(v_{21}) + v_{22}],$$

so that

$$B_1 = b_\omega = \begin{pmatrix} 0 & \partial & 1 \\ \partial & 0 & 0 \\ -1 & 0 & 0\end{pmatrix}. \tag{3.28}$$

It is easy to check out that there exists no Hamiltonian map of the form $\Phi(u)=u$, $\Phi(v)=v$, $\Phi(w)=w$, $\Phi(q)=\cdots$, which would satisfy (2.16) with

$$B=\begin{pmatrix} q\partial+\partial q & u\partial & v\partial & w\partial-\partial w \\ \partial u & 0 & \partial & 1 \\ \partial v & \partial & 0 & 0 \\ \partial w - w\partial & -1 & 0 & 0 \end{pmatrix}, \tag{3.29}$$

corresponding to the two-cocycle (3.28) on the Lie algebra $\mathfrak{g}\ltimes(V_1\oplus V_2)$.

We conclude this section with the following helpful observation.

Proposition 3.4: Let Φ: $C\to C_1$ be the Hamiltonian map (3.12) of Theorem 3.2. Denote by

$$\Phi_V:\ C_V=K\,[q_k^{(g|\nu)},c_i^{(g|\nu)},\lambda_j^{(g|\nu)}]\to C,$$

$$\Phi_{V^*}:\ C_{V^*}=K\,[q_k^{(g|\nu)},c_i^{(g|\nu)},\gamma_j^{(g|\nu)}]\to C,$$

the natural inclusions, where C_V and C_{V^*} serve the Lie algebras $\mathfrak{g}\ltimes(W\oplus V)$ and $\mathfrak{g}\ltimes(W\oplus V^*)$, respectively. Then the compositions $\Phi\Phi_V$: $C_V\to C_1$ and $\Phi\Phi_{V^*}$: $C_{V^*}\to C_1$ are Hamiltonian maps.

Proof: It is enough to show that Φ_V and Φ_{V^*} are Hamiltonian, which is evident from the formula (2.16): the contribution of the symplectic two-cocycle part (3.3) of the Hamiltonian matrix B in C vanishes, while on the remaining Hamiltonian matrices the map Φ_V (resp. Φ_{V^*}) is generated by the Lie algebra monomorphism

$$\mathfrak{g}\ltimes(W\oplus V)\to\mathfrak{g}\ltimes(W\oplus V^*\oplus V)$$

$$[\text{resp. } \mathfrak{g}\ltimes(W\oplus V)\to\mathfrak{g}\ltimes(W\oplus V^*\oplus V)].$$ ■

IV. APPLICATIONS TO ^{4}He

I first show that for the case of the nonrotating ^{4}He, the Poisson bracket (1.1) and the Hamiltonian map (1.5), are particular instances of Theorem 3.2.

Let us take $\mathfrak{g}=D_n$, $W=(\Lambda^0)^{m+1}$, $V=\Lambda^0$. Then $V^*=\Lambda^n$ by Proposition 3.5 in Ref. 3. The action of D_n on Λ^0 is given by $\rho(X)(u)=\Sigma X_k u_{,k}$ for $X=\Sigma X_k\partial_k$. By (3.9),

$$v\rho(X)(u)=\sum v_k X_k u_{,k}=X^t(u\nabla v),$$

so that

$$(u\nabla v)_k=vu_{,k}.$$

Therefore, by (3.10) and (3.11),

$$\Phi_k^1=\sum_{i=1}^{m+1} b_i a_{i,k},\quad \Phi_k^2=\psi\varphi_{,k},$$

and the map Φ in (3.12) becomes

$$\Phi(q_k)=\sum b_i a_{i,k}+\psi\varphi_{,k},$$
$$\Phi(c_i)=b_i,\quad \Phi(\gamma)=\varphi,\quad \Phi(\lambda)=\psi. \tag{4.1}$$

The Hamiltonian matrix $B=B^l+b_\omega$, for $\bar{\mathfrak{g}}=\mathfrak{g}\ltimes(W\oplus V^*\oplus V)$, equals

$$\begin{pmatrix} q_l\partial_k+\partial_l q_k & c_i\partial_k & -\gamma_{,k} & \lambda\partial_k \\ \partial_l c_i & 0 & 0 & 0 \\ \gamma_{,l} & 0 & 0 & 1 \\ \partial_l\lambda & 0 & -1 & 0 \end{pmatrix}, \tag{4.2}$$

so that Φ applied to (4.2) is the matrix (3.7). We see that there is an "invariant submanifold" $\{c_2=0,\ldots,c_{m+1}=0\}$ for (4.2). In other words, set $W=W_1\oplus W_2=\Lambda^0\oplus(\Lambda^0)^m$. Then the inclusion $W_1\to W$, $w_1\mapsto(w_1\oplus 0)$ generates a Lie algebra homomorphism

$$\mathfrak{g}\ltimes(W_1\oplus V^*\oplus V)\to\mathfrak{g}\ltimes(W\oplus V^*\oplus V).$$

The corresponding Hamiltonian map (also an inclusion)

$$\Phi':\ C'=K\,[q_k^{(g|\nu)},c_1^{(g|\nu)},\varphi^{(g|\nu)},\psi^{(g|\nu)}]$$
$$\to C=K\,[q_k^{(g|\nu)},c_i^{(g|\nu)},\varphi^{(g|\nu)},\psi^{(g|\nu)}],$$

composed with the Hamiltonian map (4.1), Φ: $C\to C_1$, provides a Hamiltonian map $\Phi''=\Phi\Phi'$: $C'\to C_1$, of the form

$$\Phi''(q_k)=\sum b_i a_{i,k}+\psi\varphi_{,k},$$
$$\Phi''(c_1)=b_1,\quad \Phi''(\gamma)=\varphi,\quad \Phi''(\lambda)=\psi. \tag{4.3}$$

The Hamiltonian matrix in C' is the same as (4.2), with only one c_1 instead of $c_1,\ldots,c_{m+1}$ present. This matrix is exactly the one producing the Poisson bracket (1.1), when the following identifications are made:

$$M_k=q_k,\quad \rho=\lambda,\quad \alpha=\gamma,\quad \sigma=c_1, \tag{4.4}$$

while the map (1.5) is just (4.3) (more precisely, dual to it), with the additional identification

$$\rho=\psi,\quad \alpha=\varphi,\quad \sigma=b_1,\quad \beta=a_1,$$
$$f_s=b_{s+1},\quad \gamma_s=a_{s+1}. \tag{4.5}$$

We now turn to the rotating ^{4}He (see Ref. 5). The Poisson bracket in this case [formula (14) in Ref. 2)] is

$$\{H,F\}\sim\left\{\frac{\delta F}{\delta M_k}\left[(M_l\partial_k+\partial_l M_k)\left(\frac{\delta H}{\delta M_l}\right)+\rho\partial_k\left(\frac{\delta H}{\delta\rho}\right)-\alpha_{,k}\frac{\delta H}{\delta\alpha}\right.\right.$$
$$\left.\left.+(\partial_l A_k-A_{l,k})\left(\frac{\delta H}{\delta a_l}\right)\right]+\left[\frac{\delta F}{\delta\rho}\partial_l\rho+\frac{\delta F}{\delta\alpha}\alpha_{,l}+\frac{\delta F}{\delta a_k}(A_{k,l}+A_l\partial_k)\right]\left(\frac{\delta H}{\delta M_l}\right)\right\} \tag{4.6a}$$
$$+\left(\frac{\delta F}{\delta\rho}\frac{\delta H}{\delta\alpha}-\frac{\delta F}{\delta\alpha}\frac{\delta H}{\delta\rho}\right) \tag{4.6b}$$
$$+\frac{\delta F}{\delta P_k}\left[(P_l\partial_k+\partial_l P_k)\left(\frac{\delta H}{\delta P_l}\right)+\sigma\partial_k\left(\frac{\delta H}{\delta\sigma}\right)\right]+\frac{\delta F}{\delta\sigma}\partial_l\sigma\left(\frac{\delta H}{\delta P_l}\right), \tag{4.6c}$$

where new notations, in addition to the nonrotating ^{4}He case, are $\mathbf{P}$ is the relative normal momentum density, and $\mathbf{A}$ is the vorticial part of the superfluid velocity $\mathbf{v}^s$, which, for rotating ^{4}He, is not curlfree anymore: $\mathbf{v}^s = \nabla\alpha - \mathbf{A}$.

The bracket (4.6) splits off in two separate brackets: [(4.6a,b) and (4.6c)]. The bracket (4.6c) is the natural bracket on the semidirect product Lie algebra

$$\bar{\mathfrak{g}}_2 = D_n \circledS \Lambda^0. \tag{4.7}$$

The bracket (4.6a) and (4.6b) is of the form $B = B^l(\bar{\mathfrak{g}}_1) + b$, where $B^l(\bar{\mathfrak{g}}_1)$, given by (4.6a), is naturally associated with the semidirect product Lie algebra

$$\bar{\mathfrak{g}}_1 = D_n \circledS (\Lambda^{n-1} \oplus \Lambda^0 \oplus \Lambda^n), \tag{4.8}$$

while b, given by (4.6b), is the symplectic two-cocycle on the $\Lambda^0 \oplus \Lambda^n$ part of $\bar{\mathfrak{g}}_1$. [Notice the transposition of Λ^n and Λ^0 in (4.8) in contrast to the nonrotating ^{4}He case (1.2), which is responsible for nonrotating (1.1b) and rotating two-cocycles (4.6b) having *opposite signs.*]

The bracket (4.6) was obtained in Ref. 2 by a direct mathematical computation of the following sort (in the present notation): The map Φ, given by the formulas,

$$\Phi(P_k) = -\sigma\beta_{,k} - f\Gamma_{,k}, \quad \Phi(\sigma) = \sigma, \tag{4.9}$$

$$\Phi(M_k) = -\rho\alpha_{,k} + \sum [-d_j A_{j,k} + (d_j A_k)_{,j}],$$

$$\Phi(\rho) = \rho, \quad \Phi(\alpha) = \alpha, \quad \Phi(A_k) = A_k, \tag{4.10}$$

is Hamiltonian between (4.6) and the symplectic Poisson bracket in space with canonical pairs of variables $(\sigma;\beta)$, $(f;\Gamma)$, $(\rho;\alpha)$, and (d_k,A_k). Let us see that this fact is a particular instance of Theorem 3.2. First, the maps (4.9) and (4.10) split off, with (4.9) associated with the bracket (4.6c), and this part is covered by Proposition 3.4 applied to the cocycleless case $D_n \circledS \Lambda^0$ (see the remark in Sec. III) of Theorem 3.2, followed by ignoring the variable f in the same fashion as the variables $c_2,\ldots,c_{m+1}$ were purged from the matrix (4.2). Formula (4.10) can be gotten from the following computations. For $W = \Lambda^{n-1}$, the action of $X = \Sigma X_k \partial_k$ on $\omega = \Sigma\omega_j(\partial_j \lrcorner\, d^n x)$, where $d^n x = dx_1 \wedge \cdots \wedge dx_n$, is

$$X(\omega)_i = \sum (X_k \omega_{i,k} + \omega_i X_{k,k} - \omega_k X_{i,k}). \tag{4.11}$$

Hence, by (3.9),

$$v^t X(u) = \sum v_i (X_k u_{i,k} + u_i X_{k,k} - u_k X_{i,k})$$

$$\sim \sum X_k [v_i u_{i,k} - (u_i v_i)_{,k} + (v_k u_i)_{,i}],$$

so that

$$(u\nabla v)_k = \sum [-v_{i,k} u_i + (v_k u_i)_{,i}]. \tag{4.12}$$

Thus, by (3.10),

$$\Phi^1_k = (a\nabla b)_k = \sum [-b_{i,k} a_i + (b_k a_i)_{,i}]. \tag{4.13}$$

Also, for $V = \Lambda^n$,

$$X(\omega\, d^n x) = \sum (X_k \omega)_{,k}\, d^n x,$$

and hence, by (3.9),

$$v^t X(u) = \sum v(X_k u)_{,k} \sim -\sum X_k u v_{,k},$$

so that

$$(u\nabla v)_k = -u v_{,k}.$$

Thus,

$$\Phi^2_k = (\varphi\nabla\psi)_k = -\varphi\psi_{,k}. \tag{4.14}$$

Comparing now (3.12) with (4.10), we see that they coincide, provided we make the identifications

$$\begin{aligned} &M_k = q_k, \quad \varphi = \gamma = \rho, \quad \psi = \lambda = \alpha, \\ &a_k = d_k, \quad b_k = c_k = A_k. \end{aligned} \tag{4.15}$$

The sign in (4.6b) is exactly that of the symplectic two-cocycle part in (3.14).

ACKNOWLEDGMENT

This work was partially supported by the National Science Foundation and the U.S. Department of Energy.

[1] I. E. Dzyaloshinskii and G. E. Volovick, Ann. Phys. (NY) **125**, 67 (1980).
[2] D. D. Holm and B. A. Kupershmidt, Phys. Lett. A **91**, 425 (1982).
[3] B. A. Kupershmidt, J. Math. Phys. **26**, 2754 (1985).
[4] B. A. Kupershmidt, "Discrete Lax equations and differential difference calculus," in *Revue Asterisque* (Paris, 1985), Vol. 123.
[5] V. L. Lebedev and I. M. Khalatnikov, JETP Lett. **28**, 83 (1978).

Erratum: Comment on an aspect of a paper by G. Thompson [J. Math. Phys. 27, 153 (1986)]

P. G. L. Leach
Department of Applied Mathematics, The University of the Witwatersrand, 1 Jan Smuts Avenue, Johannesburg, South Africa 2001

(Received 17 April 1986; accepted for publication 23 May 1986)

A mistake in the solution of an equation in the paper "Comment on an aspect of a paper by G. Thompson"[1] is rectified.

The correct solution to Eq. (2.11) in Ref. 1 is

$$V(\eta,\zeta) = K(\eta)e^{-2\zeta} + L(\eta)e^{-3\zeta} + M(\zeta)\,, \tag{1}$$

and (2.20) is now

$$(a + b'e^{-\zeta})(K'e^{-2\zeta} + L'e^{-3\zeta}) + be^{-\zeta}(-2Ke^{-2\zeta} - 3Le^{-3\zeta} + M') = 0\,, \tag{2}$$

where

$$K' = \tfrac{1}{3}\alpha', \quad L' = \tfrac{1}{3}(b'' + b)\,. \tag{3}$$

The analysis of (2) separates into two cases, $b \equiv 0$ and $b \not\equiv 0$.

Case I: $b \equiv 0$. From (2) and (3) we have

$$L' = 0, \quad aK' = 0, \quad aa' = 0\,, \tag{4}$$

whence

$$L = L_0, \quad K = K_0, \quad A = A_0\,. \tag{5}$$

Hence for the Hamiltonian

$$\begin{aligned} H &= \tfrac{1}{2}e^{-2\eta}(p_\eta^2 + p_\zeta^2) + L_0e^{-2\zeta} + K_0e^{-3\zeta} + M(\zeta) \\ &= \tfrac{1}{2}e^{-2\zeta}(p_\eta^2 + p_\zeta^2) + \mathscr{M}(\zeta)\,, \end{aligned} \tag{6}$$

we obtain the first integral

$$I = p_\eta^3 + A_0 p_\eta\,. \tag{7}$$

The potential is just of the form $V(r)$, where r is the radial variable and the first integral is not truly cubic, but is a cubic function of a first integral linear in the momentum p_η, the conserved angular momentum.

Case II: $b \not\equiv 0$. Rewriting (2) as

$$M'(\zeta) = -\frac{aK'}{b}e^{-\zeta} + \left(2K - \frac{aL'}{b} - \frac{b'K}{b}\right)e^{-2\zeta} + \left(3L - \frac{b'L'}{b}\right)e^{-3\zeta}\,, \tag{8}$$

it is evident that the partial derivative of the right-hand side with respect to η is zero for each of the coefficients of the three linearly independent functions of ζ. Thus, taking (3) into account,

$$(a\,a'/b)' = 0\,, \tag{9}$$

$$2a' - [(a/b)(b'' + b) + b'a'/b]' = 0\,, \tag{10}$$

$$3(b'' + b) - [(b'/b)(b'' + b)]' = 0\,. \tag{11}$$

Integrating (9),

$$aa'/b = N\,, \tag{12}$$

where N is a constant which may be zero.

We dispose of the case $N \neq 0$. Three successive quadratures of (10) yield

$$a'^2 = \tfrac{1}{6}a^2 + Ba + \; + D/a^2\,, \tag{13}$$

in which B, C, and D are arbitrary constants. From (12)

$$b = (1/N)(\tfrac{1}{6}a^4 + Ba^3 + Ca + D)^{1/2}\,. \tag{14}$$

However, b does not satisfy (11) and so we conclude that the assumption $N \neq 0$ is invalid and that $a = 0$ or $a' = 0$.

Case II (i): $a' \equiv 0$. From (10)

$$b'' + b = Cb\,, \tag{15}$$

where C is some constant. Substituting (15) into (11),

$$C(b'' - 3b) = 0\,, \tag{16}$$

we see that either $C = 0$ or 4 and we obtain the two possible sets of solutions:

$$\begin{aligned} &a(\eta) = K_0, \quad K(\eta) = K_0, \quad L(\eta) = L_0\,, \\ &b(\eta) = B_1 \sin\eta + B_2 \cos\eta\,, \end{aligned} \tag{17}$$

corresponding to $C = 0$; and

$$\begin{aligned} &a(\eta) = A_0, \quad (K\eta) = K_0\,, \\ &b(\eta) = B_1 \sinh \eta\sqrt{3} + B_2 \cosh \eta\sqrt{3}\,, \\ &L(\eta) = (4/3\sqrt{3})(B_1 \cosh \eta\sqrt{3} + B_2 \sinh \eta\sqrt{3}) + L_0\,, \end{aligned} \tag{18}$$

corresponding to $C = 4$. In the first case the potential is zero, i.e., we have the free particle and the first integral

$$I = p_\eta^3 + \{A_0 + (B_1 \cos\eta - B_2 \sin\eta)e^{-\zeta}\}p_\eta + (B_1 \sin\eta + B_2 \cos\eta)e^{-\zeta}p_\zeta \tag{19}$$

is not a first integral that is truly cubic in p_η since p_η is itself a first integral.

Turning to the second case the potential is

$$V(\eta,\zeta) = \tfrac{2}{3}A_0e^{-2\zeta} + (4/3\sqrt{3})(B_1 \cosh \eta\sqrt{3} + B_2 \sinh \eta\sqrt{3})e^{-3\zeta} \tag{20}$$

and the first integral is

$$I = p_\eta^3 + \{A_0 + \sqrt{3}(B_1 \cosh \eta\sqrt{3} + B_2 \sinh \eta\sqrt{3})e^{-\zeta}\}p_\eta + (B_1 \sinh \eta\sqrt{3} + B_2 \cosh \eta\sqrt{3})e^{-\zeta}p_\zeta\,. \tag{21}$$

In this case we have, for B_1 and B_2 not both zero, a first integral that is truly cubic in momenta.

Case II (ii): $a \equiv 0$. Equations (9)–(11) now reduce to (11) only. Apart from the solutions for $b(\eta)$ given in (17) and (18) we have not been able to obtain a general solution for (11). The only explicit results we can report are (19)–(21) with A_0 set at zero.

In the analysis of the problem of the existence of a first integral of the form

$$I = p_\eta^4 + A(\eta,\zeta)p_\eta^2 + B(\eta,\zeta)p_\eta p_\zeta + C(\eta,\zeta)p_\zeta^2 + D(\eta,\zeta)\,, \tag{22}$$

a similar mistake was made and the solution of (3.2) in Ref. 1 should read

$$V(\eta,\zeta) = K(\eta)e^{-2\zeta} + L(\eta)e^{-3\zeta} + M(\eta)e^{-4\zeta} + N(\zeta)\,. \tag{23}$$

Applying the Poisson bracket condition on I we find that

$$\begin{aligned}
&C = c(\eta)e^{-2\zeta}, \quad B = b(\eta)e^{-\zeta} + c'e^{-2\zeta},\\
&A = a(\eta) + b'e^{-\zeta} + (\tfrac{1}{2}c'' + c)e^{-2\zeta},\\
&M' = \tfrac{1}{4}c', \quad L' = \tfrac{1}{4}b\,,\\
&K' = \tfrac{1}{4}(\tfrac{1}{2}c''' + c'), \quad b'' = 0, \quad a' = 0\,,
\end{aligned} \tag{24}$$

and the function $D(\eta,\zeta)$ is determined by

$$\frac{\partial D}{\partial \eta} = 2A(K' + L'e^{-\zeta} + M'e^{-2\zeta}) - B(2K + 3Le^{-\zeta} + 4Me^{-2\zeta} - N'e^{2\zeta})\,, \tag{25}$$

$$\frac{\partial D}{\partial \zeta} = B(K' + L'e^{-\zeta} + M'e^{-2\zeta}) - 2C(2K + 3Le^{-\zeta} + 4Me^{-2\zeta} - N'e^{2\zeta})\,. \tag{26}$$

Writing $W = N'\exp(2\zeta)$, the consistency requirement on (25) and (26) is

$$\begin{aligned}
&W'(be^{-\zeta} + c'e^{-2\zeta}) - W(be^{-\zeta} + 4c'e^{-2\zeta})\\
&\quad + 2\{-b'e^{-\zeta} - (c'' + 2c)e^{-2\zeta}\}\\
&\quad \times\{K' + L'e^{-\zeta} + M'e^{-2\zeta}\}\\
&\quad + 2\{a + b'e^{-\zeta} + (\tfrac{1}{2}c'' + c)e^{-2\zeta}\}\\
&\quad \times\{-L'e^{-\zeta} - 2M'e^{-2\zeta}\}\\
&\quad + (be^{-\zeta} + 4c'e^{-2\zeta})(2K + 3Le^{-\zeta} + 4Me^{-2\zeta})\\
&\quad + (be^{-\zeta} + c'e^{-2\zeta})(3Le^{-\zeta} + 8Me^{-2\zeta})\\
&\quad - (b'e^{-\zeta} + c''e^{-2\zeta})(K' + L'e^{-\zeta} + M'e^{-2\zeta})\\
&\quad - (be^{-\zeta} + c'e^{-2\zeta})(K'' + L''e^{-\zeta} + M''e^{-2\zeta})\\
&\quad + 2ce^{-2\zeta}(2K' + 3L'e^{-\zeta} + 4M'e^{-2\zeta}) = 0\,.
\end{aligned} \tag{27}$$

For (27) to be a differential equation for W, it is necessary for $b \neq 0$ and $c' \neq 0$ and for its nonhomogenous term when divided by the coefficient of W' or W to be η-free. In both cases a contradiction occurs in that the result is $b = 0 = c'$. Hence W and so N is an arbitrary function of ζ.

Substituting in (25) and (26) we find

$$D(\eta,\zeta) = 2C_0K_0e^{-2\zeta} + 2C_0L_0e^{-\zeta} + 2C_0N(\zeta) \tag{28}$$

and that

$$I = p_\eta^4 + A_0 p_\eta^2 + 2C_0H \tag{29}$$

is a first integral for the Hamiltonian

$$\begin{aligned}
H &= \tfrac{1}{2}e^{-2\zeta}(p_\eta^2 + p_\zeta^2) + K_0e^{-2\zeta}\\
&\quad + L_0e^{-3\zeta} + M_0e^{-4\zeta} + N(\zeta)\\
&= \tfrac{1}{2}e^{-2\zeta}(p_\zeta^2 + p_\eta^2) + \mathcal{N}(\zeta)\,.
\end{aligned} \tag{30}$$

As p_η is itself a first integral I is not truly a first integral quartic in the momenta.

ACKNOWLEDGMENTS

I thank Dr. Tanaji Sen of The State University of New York at Stony Brook for informing me of the error in Ref. 1 and G. C. Hillebrand for analyzing (27) using REDUCE.

[1] P. G. L. Leach, J. Math. Phys. 27, 153 (1986).