Totally symmetric irreducible representations of the group SO(6) in the principal SO(3) subgroup basis

J. Vanthournout, J. Van der Jeugt,[a] H. De Meyer,[b] and G. Vanden Berghe
Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit-Gent, Krijgslaan 281-gebouw S9, B9000 Gent, Belgium

(Received 3 June 1986; accepted for publication 22 July 1987)

Explicit matrix elements are found for the generators of the group SO(6) in an arbitrary totally symmetric irreducible representation, using the physical principal SO(3) subgroup in the chain $SO(6)\supset SO(5)\supset SO(3)$. The internal one missing label problem is solved through the definition of intrinsic states associated to the $SU(2)\times SU(2)$ subgroup in the chain $SO(6)\supset SO(5)\supset SU(2)\times SU(2)$ and out of which is projected a complete set of states in the physical basis by integrations over the physical rotation group manifold. The matrix elements of the SO(6) generators in the $SU(2)\times SU(2)$ basis are themselves obtained by the intermediate use of an $SU(2)\times SU(2)\times U(1)$ basis, the latter group being a subgroup of SO(6) but not of SO(5).

I. INTRODUCTION

A well-known and remarkably successful model for describing even–even nuclei is the standard interacting boson model.[1–3] The basic constituents of this model are an s and d boson, carrying angular momentum 0 and 2, respectively. The bilinear forms in the associated boson creation operators ($s^\dagger$, $d_\mu^\dagger$: $\mu=-2,-1,\ldots,2$) and boson annihilation operators (s, d_μ: $\mu=-2,-1,\ldots,2$) generate a U(6) Lie group. The principle of boson number conservation permits to restrict the group to SU(6) or even to its local contents described by the SU(6) Lie algebra. The interacting boson model thus allows for three-types of so-called dynamical symmetries, mathematically connected to the three Lie algebra inclusion chains $SU(6)\supset SU(3)\supset SO(3)$, $SU(6)\supset SO(6)\supset SO(5)\supset SO(3)$, and $SU(6)\supset SU(5)\supset SO(5)\supset SO(3)$. Each of the chains ends at the physical SO(3) Lie subalgebra of SU(6). Hence, the N-boson states belonging to the totally symmetric irreducible representation (irrep) [N0000] of SU(6) are partly labeled by the angular momentum l and its projection m along a fixed axis, and partly by the representation labels of the subalgebra(s) in the chain considered. In general still one label is missing in order to distinguish completely between the SU(6) states.

To each of the dynamical symmetries corresponds an unperturbed Hamiltonian[1–3] which is expressible in terms of the second degree Casimir operators of the algebras which occur in the associated symmetry chain. Although the theoretical predictions resulting from this simplest kind of model Hamiltonian do fit the experimental data rather well, there is nowadays a tendency to consider higher degree boson interaction terms into the Hamiltonian as well. The interaction terms which are frequently discussed[4–6] usually break the existing dynamical symmetry. Recently it has been argued by some of the present authors[7] that from the mathematical point of view there is no *a priori* reason to destroy the symmetry. Indeed, it suffices to select the higher degree interaction terms out of the set of SO(3) scalars belonging to the enveloping algebra of the dynamical symmetry subalgebra. In fact, one can be even more restrictive by choosing the interactions out of the so-called integrity basis for SO(3) scalars.[8]

This program of systematic, complete, and symmetry conserving interaction term generation has been successfully carried out in the SU(3) symmetry limit by some of us.[7] More in detail, we have obtained, apart from the SU(3) and SO(3) invariants, two functionally independent SO(3) scalars in the SU(3) enveloping algebra, which were added to the unperturbed Hamiltonian and both proved to bear physical relevance, since they account for theoretical results which fit much better the experimental measurements. Also, we have been able to establish a simple algorithm to evaluate matrix elements of these operators in any totally symmetric SU(6) irrep.[9–11]

Some years ago, three of the present authors have made in a series of papers a thorough study of the classification of nuclear quadrupole and octupole phonon states.[12,13] One of the particular aspects therein has been the construction of an SO(3) scalar of fourth degree in the SO(5) enveloping algebra. Also large parts of its spectrum have been derived in closed form. In the scope of extending the interacting boson model while preserving the symmetry, this operator is clearly relevant both in the SO(6) and the SU(5) limits. It is, nevertheless, in these cases more natural to ask for SO(3) scalars which preserve the SO(6) or SU(5) symmetry but eventually break the SO(5) symmetry. In fact, it has been proved recently by one of us[14] that in both limits there exist in the integrity basis of the corresponding enveloping algebra two functionally independent SO(3) scalars of third degree in the generators. Hence, if one wishes to treat the SO(6) and SU(5) dynamical symmetries in complete analogy with the SU(3) limit, it is of primordial interest to construct first these operators and to establish a method for finding their matrix elements in the corresponding physically labeled state basis. The present paper accounts for a first decisive step towards the achievement of that goal.

[a] Senior Research Assistant N. F. W. O., Belgium.
[b] Senior Research Associate, N. F. W. O., Belgium.

 0022-2488/87/112529-11$02.50

The SO(3) scalar operators concerned can be viewed as Racah-coupled generators which are themselves equivalent to SO(3) tensor operators. Hence, it suffices at first instance to compute the matrix elements of the SO(6) or SU(5) generators in the physical basis which exhibits the symmetry under consideration. It is the aim of the present work to carry out this task for the SO(6) limit. The SU(5) limit will be treated in a forthcoming publication.

Our analysis has been inspired and influenced by the earlier work of Kemmer *et al.*[15,16] in which the matrix elements of the SO(5) (subalgebra) generators have been established by means of closed formulas. The analogy, however, is far from being trivial, as we shall have to introduce besides the SU(2)×SU(2) subalgebra which they considered, another subalgebra of the type SU(2)×SU(2)×U(1) which is, however, not contained in SO(5).

The paper is outlined as follows. In Sec. II we describe the SO(6) algebra and its irreducible representations (irreps) in the physical SO(3) basis. Section III is concerned with a similar discussion with respect to the SU(2)×SU(2) basis. The already mentioned SU(2)×SU(2)×U(1) subalgebra is treated in Sec. IV. In this basis we shall at first succeed in deriving the SO(6) generator matrix elements. We next return to the physical basis going over the SU(2)×SU(2) basis in Sec. V and arriving at our final results in Sec. VI.

II. SO(6) IN THE PHYSICAL SO(3) BASIS

It is a well-known fact that the SO(6) irreps reduce into irreps of the physical SO(3) Lie subalgebra according to the physical reduction chain SO(6) ⊃ SO(5) ⊃ SO(3).[17] In particular, it is verified from standard tables[18,19] that the 15-dimensional SO(6) irrep [0,1,1] decomposes throughout that chain in the SO(3) irreps [1], [2], and [3], whereby $2k+1$ is the dimension of the irrep $[k]$. Hence, in the SO(3) basis the SO(6) Lie algebra is generated by the SO(3) basis elements $l_0, l_{\pm 1}$ together with the components p_μ $(\mu=-2,-1,...,2)$ of a five-dimensional irreducible SO(3) tensor operator and the components q_μ $(\mu=-3,-2,...,3)$ of a seven-dimensional SO(3) tensor operator. The l and q operators form a basis of the ten-dimensional SO(5) Lie subalgebra in the chain. Since p and q are SO(3) tensors of rank 2 and 3, respectively, the following commutation properties hold by definition:

$$[l_{-1},l_{+1}]=l_0,\quad [l_0,l_{\pm 1}]=\pm l_{\pm 1},\tag{2.1}$$

$$[l_0,p_\mu]=\mu p_\mu,$$

$$[l_{\pm 1},p_\mu]=\mp\sqrt{(2\mp\mu)(3\pm\mu)/2}\,p_{\mu\pm 1}\quad(\mu=-2,-1,...,2),\tag{2.2}$$

$$[l_0,q_\mu]=\mu q_\mu,$$

$$[l_{\pm 1},q_\mu]=\mp\sqrt{(3\mp\mu)(4\mp\mu)/2}\,q_{\mu\pm 1}\quad(\mu=-3,-2,...,3).\tag{2.3}$$

The remaining commutators can be derived by expressing the SO(3) tensors $l,p,$ and q in terms of the canonical SO(3) tensor operators $G^k_\kappa(l,l')$ which map the $(2l'+1)$-dimensional space specified by l' into a $(2l+1)$-dimensional space and which are unambiguously defined by their reduced matrix elements, i.e.,[20]

$$\langle l'''\|G^k(l,l')\|l''\rangle=[2k+1]^{1/2}\delta_{l''l}\delta_{l'l''}.\tag{2.4}$$

These canonical SO(3) tensor operators are known to satisfy the commutation property

$$\begin{aligned}[G^{k_1}_{\kappa_1}(l_1,l_2),G^{k_2}_{\kappa_2}(l_3,l_4)]
&=\sum_{k_3,\kappa_3}[(2k_1+1)(2k_2+1)(2k_3+1)]^{1/2}\\
&\times\begin{pmatrix}k_1&k_2&k_3\\ \kappa_1&\kappa_1&-\kappa_3\end{pmatrix}(-1)^{2l_4+l_3-l_2-\kappa_3}\\
&\times\Bigg(\delta_{l_2l_3}(-1)^{k_1+k_2+k_3+l_1+l_2+l_3+l_4}\begin{Bmatrix}k_1&k_2&k_3\\ l_4&l_1&l_3\end{Bmatrix}\\
&\times G^{k_3}_{\kappa_3}(l_1,l_4)-\delta_{l_1l_4}\begin{Bmatrix}k_1&k_2&k_3\\ l_3&l_2&l_1\end{Bmatrix}G^{k_3}_{\kappa_3}(l_3,l_2)\Bigg).\end{aligned}\tag{2.5}$$

For the case under study, since the lowest dimensional SO(6) irrep decomposes into the SO(3) irreps [0] and [2], the representation space is specified by $l=0$ and $l=2$. Moreover, it has already been established that[20–22]

$$\begin{aligned}l_\mu&=\sqrt{10}\,G^{(1)}_\mu(2,2)\quad(\mu=-1,0,1),\\
p_\mu&=G^{(2)}_\mu(2,0)+G^{(2)}_\mu(0,2)\quad(\mu=-2,-1,...,2),\\
q_\mu&=G^{(3)}_\mu(2,2)\quad(\mu=-3,-2,...,3).\end{aligned}\tag{2.6}$$

The commutators amongst p and q operators now follow by combining (2.6) and (2.5), whereas the already obtained commutators (2.1)–(2.3) can be verified again.

In the present work we are only concerned with SO(6) as one of the dynamical symmetry groups of the SU(6) interacting boson model. Hence, we need to construct for any totally symmetric SU(6) irrep $[N,0,0,0,0]$ a nondegenerate state basis which exhibits the SO(6) and the physical SO(3) symmetries. On account of the reduction formula

$$\mathrm{SU}(6)\to\mathrm{SO}(6):\ [N,0,0,0,0]\to\sum_\sigma[\sigma,0,0],\quad\text{with } \sigma=N,N-2,N-4,...,1\text{ or }0,\tag{2.7}$$

we can consider the SU(6) symmetric irreps to be fully reduced to SO(6) so that we can confine ourselves to the problem of the decomposition of the totally symmetric irreps of SO(6) with respect to SO(3). Moreover, in the reduction of SO(6) into SO(5) these irreps also decompose without degeneracy into a sum of totally symmetric SO(5) irreps, namely,

$$\mathrm{SO}(6)\to\mathrm{SO}(5):\ [\sigma,0,0]\to\sum_\tau[\tau,0],\quad\text{with } \tau=\sigma,\sigma-1,\sigma-2,...,0.\tag{2.8}$$

It is well known that in the further reduction of the symmetric SO(5) irreps into SO(3) irreps one additional label is required.[15,23,24] From what we have learned out of our treatment of the nuclear quadrupole vibrations[12,13] the fourth degree SO(5) scalar in the SO(5) enveloping algebra could be used as a label generating operator. However, because its spectrum cannot be easily described, it is in the light of the present calculations more convenient to choose as the extra label the one which is usually considered in the interacting boson model[17] and which is obtained in a manner closely analogous to that of Elliott for SU(3),[21] i.e., the label ν (sometimes denoted by ν_Δ) which takes on the values

$$\nu = 0,1,2,\ldots,[\tau/3], \tag{2.9}$$

whereas for a given ν value the SO(3) contents is specified by

$$l = 2(\tau - 3\nu), 2(\tau - 3\nu) - 2,$$
$$2(\tau - 3\nu) - 3,\ldots,\tau - 3\nu + 1, \tau - 3\nu. \tag{2.10}$$

Hence, in any SO(6) symmetric irrep $[\sigma,0,0]$ the states which constitute a physical basis are labeled as

$$|\sigma,\tau,\nu,l,m\rangle \quad (m = -l, -l+1,\ldots,l), \tag{2.11}$$

where one has to take into account the restrictions imposed by (2.7)–(2.10). It should be remarked that the states are orthogonal in all labels except for the ν label. More explicitly,

$$\langle \sigma\tau'\nu'l'm'|\sigma\tau\nu lm\rangle = A_l^\tau(\nu',\nu)\delta_{\tau'\tau}\delta_{l'l}\delta_{m'm}. \tag{2.12}$$

Several equivalent formulas which permit to evaluate the overlap integrals $A_l^\tau(\nu',\nu)$ have been derived by Williams and Pursey.[16] For the sake of self-containedness we mention here one of the expressions

$$\begin{aligned} A_l^\tau(\nu',\nu) = {}& 2^{\nu'-\nu}[(2l+1)(3\nu'-3\nu)!]^{-1}[(\tau-\nu)!(\tau-\nu')!\nu!\nu'!(l+3\nu'-\tau)! \\ & \times (l+\tau-3\nu)!/(l+\tau-3\nu')!(l+3\nu-\tau)!]^{1/2}\sum_{\alpha,\beta}(-4)^{\nu+\alpha-\beta}(3\nu'-3\beta+\alpha)! \\ & \times (2\tau-2\nu-2\nu'+2\beta)![(\tau-\nu-\nu'+\beta-\alpha)!(\nu'-\beta)!(\nu-\beta)!\alpha!\beta!(2\tau+\nu'-2\nu+\alpha-\beta+1)!]^{-1} \\ & \times {}_3F_2(\tau-3\nu-l,\tau+l-3\nu+1,3\nu'-3\beta+\alpha+1;3\nu'-3\nu+1,2\tau+\nu'-2\nu-\beta+\alpha+2;1). \end{aligned} \tag{2.13}$$

Herein $_3F_2$ represents a generalized hypergeometric function which on the rhs of (2.13) always reduces into a polynomial.

It is our final aim to obtain the matrix elements of the SO(6) generators in the physical but nonorthonormalized basis (2.11). In fact, part of the work is either trivial or has been carried out already previously. Indeed, concerning the matrix elements of the l operator and taking into account (2.12), we immediately obtain that

$$\langle \sigma,\tau',\nu',l'\|\mathbf{l}\|\sigma,\tau,\nu,l\rangle = \delta_{\tau'\tau}\delta_{l'l}[l(l+1)(2l+1)]^{1/2}A_l^\tau(\nu',\nu), \tag{2.14}$$

where we have preferred to mention the SO(3) reduced matrix elements instead. Furthermore, the reduced matrix elements of the SO(5) generators q_μ ($\mu = -3,-2,\ldots,3$) have been found by Williams and Pursey.[16] They read

$$\begin{aligned} &\langle \sigma\,\tau'\,\nu'\,l'\|q\|\sigma\,\tau\,\nu\,l\rangle \\ &= \delta_{\tau'\tau}(2l'+1)^{1/2}\{[-(5\nu(\tau-\nu+1))^{1/2}\langle lK\,3\,3|l'\,K+3\rangle \\ &\quad -(5\nu l'(l'+1)/3(\tau-\nu+1))^{1/2}\langle l'\,K+3\,1-1|l'\,K+2\rangle\langle lK\,3\,2|l'\,K+2\rangle]A_{l'}^\tau(\nu',\nu-1) \\ &\quad +(5(\tau-\nu)(\nu+1))^{1/2}\langle lK\,3\,-3|l'\,K-3\rangle A_{l'}^\tau(\nu',\nu+1) \\ &\quad +[(3l'(l'+1)/2)^{1/2}\langle l'\,K\,1\,1|l'K+1\rangle\langle lK\,3\,1|l'\,K+1\rangle \\ &\quad -(2l'(l'+1)/3)^{1/2}\langle l'\,K\,1\,-1|l'\,K-1\rangle\langle lK\,3-1|l'\,K-1\rangle \\ &\quad -(2\tau-\nu)\langle lK\,3\,0|l'K\rangle]A_{l'}^\tau(\nu',\nu)\} \quad (K=\tau-3\nu), \end{aligned} \tag{2.15}$$

wherein $\langle l\,m\,l'\,m'|l''\,m+m'\rangle$ denotes a Clebsch–Gordan coefficient. All that remains to be done is the calculation of the reduced matrix elements of the p-type generators. Of course, we can predict that the matrices will no longer be diagonal with respect to the SO(5) label τ, as it is the case with the matrix (2.15). Since Kemmer *et al.*[15,16] provided us with a technique to project the physically labeled states (2.11) out of another basis of states described by means of SU(2)×SU(2) representation labels, it seems straightforward, as a start for solving the present problem, to consider that particular subalgebra too.

III. SO(6) IN THE SU(2)×SU(2) BASIS

The SO(5) subalgebra of SO(6) contains itself an SU(2)×SU(2) Lie subalgebra in which, however, the physical SO(3) is not included. Nevertheless, the Hill–Wheeler projection technique will enable us to go over from the SU(2)×SU(2) basis into the physical basis. From the standard reduction tables one can see that the SO(5) generators decompose into the SU(2)×SU(2) generators together with an SU(2)×SU(2) spinor–spinor representation $T^{(1/2,1/2)}$. Besides, the p operators which also constitute a five-dimensional SO(5) tensor representation can be rearranged to form the components of both another SU(2)×SU(2) bispinor $U^{(1/2,1/2)}$ and an SU(2)×SU(2) scalar z_0. Denoting further the generators of the first SU(2) group by s_μ ($\mu = -1,0,1$) and those of the second SU(2) group by t_μ ($\mu = -1,0,1$), the following relationship between the two SO(6) bases can be established:

$$l_0 = 3s_0 + t_0,\quad l_{\pm 1} = 2t_{\pm 1} \mp \sqrt{3}T^{(1/2,1/2)}_{\pm 1/2,\mp 1/2},\quad q_0 = (1/\sqrt{10})(-s_0 + 3t_0),$$
$$q_{\pm 1} = \sqrt{\tfrac{3}{5}}\,t_{\pm 1} + (1/\sqrt{5})T^{(1/2,1/2)}_{\pm 1/2,\mp 1/2},\quad q_{\pm 2} = \pm(1/\sqrt{2})T^{(1/2,1/2)}_{\pm 1/2,\pm 1/2}, \tag{3.1}$$
$$q_{\pm 3} = s_{\pm 1},\quad p_0 = z_0,\quad p_{\pm 1} = U^{(1/2,1/2)}_{\pm 1/2,\mp 1/2},\quad p_{\pm 2} = U^{(1/2,1/2)}_{\pm 1/2,\pm 1/2}.$$

In the new SU(2)×SU(2) basis the commutation relations read

$$\begin{aligned}
&[s_0,s_{\pm 1}] = \pm s_{\pm 1},\quad [s_{-1},s_{+1}] = s_0,\quad [t_0,t_{\pm 1}] = \pm t_{\pm 1},\quad [t_{-1},t_{+1}] = t_0,\quad [s_0,T_{\alpha\beta}] = \alpha T_{\alpha\beta},\\
&[t_0,T_{\alpha\beta}] = \beta T_{\alpha\beta},\quad [s_0,U_{\alpha\beta}] = \alpha U_{\alpha\beta},\quad [t_0,U_{\alpha\beta}] = \beta U_{\alpha\beta},\quad [s_{\pm 1},T_{\mp 1/2,\beta}] = \mp(1/\sqrt{2})T_{\pm 1/2,\beta},\\
&[t_{\pm 1},T_{\alpha,\mp 1/2}] = \mp(1/\sqrt{2})T_{\alpha,\pm 1/2},\quad [s_{\pm 1},U_{\mp 1/2,\beta}] = \mp(1/\sqrt{2})U_{\pm 1/2,\beta},\quad [t_{\pm 1},U_{\alpha,\mp 1/2}] = \mp(1/\sqrt{2})U_{\alpha,\pm 1/2},\\
&[T_{1/2,\pm 1/2},T_{-1/2,\mp 1/2}] = \mp s_0 - t_0,\quad [U_{1/2,\pm 1/2},U_{-1/2,\mp 1/2}] = \pm s_0 + t_0,\\
&[T_{\pm 1/2,1/2},T_{\pm 1/2,-1/2}] = -\sqrt{2}\,s_{\pm 1},\quad [U_{\pm 1/2,1/2},U_{\pm 1/2,-1/2}] = \sqrt{2}\,s_{\pm 1},\\
&[T_{1/2,\pm 1/2},T_{-1/2,\pm 1/2}] = -\sqrt{2}\,t_{\pm 1},\quad [U_{1/2,\pm 1/2},U_{-1/2,\pm 1/2}] = \sqrt{2}\,t_{\pm 1},\\
&[T_{1/2,\pm 1/2},U_{-1/2,\mp 1/2}] = [T_{-1/2,\mp 1/2},U_{1/2,\pm 1/2}] = \mp z_0,\quad [z_0,T_{\alpha\beta}] = -U_{\alpha\beta},\\
&[z_0,U_{\alpha\beta}] = -T_{\alpha\beta}\quad (\forall\alpha,\beta\in\{\tfrac{1}{2},-\tfrac{1}{2}\})
\end{aligned} \tag{3.2}$$

All commutators that are not cited in (3.2) are zero. Also we have dropped the superscripts in T and U.

In the reduction chain SO(6)⊃SO(5)⊃SU(2)×SU(2) a totally symmetric SO(6) irrep $[\sigma,0,0]$ consecutively decomposes as prescribed by the rule (2.8) and by[15]

$$\mathrm{SO}(5)\to\mathrm{SU}(2)\times\mathrm{SU}(2):\ [\tau,0]\to\sum_{s=t}(s,t),$$

with

$$s = t = \tau/2,\tau/2-\tfrac{1}{2},\tau/2-1,\ldots,\tfrac{1}{2},0. \tag{3.3}$$

Notice that the reduction is complete, hence no labels are missing. We denote the orthonormal SO(6) basis states which make the SU(2)×SU(2) subalgebra apparent, by

$$|\sigma,\tau,s,m_s,t,m_t)$$
$$(m_s = -s,-s+1,\ldots,s;\ m_t = -t,-t+1,\ldots,t), \tag{3.4}$$

where, of course, on account of (3.3) $s=t$ for all states belonging to symmetric SO(6) irreps.

The matrix elements of the SO(5) generators, i.e., the s_μ, t_μ, and $T_{\alpha\beta}$ operators, expressed in the SU(2)×SU(2) basis (3.4) have been discussed by Kemmer *et al.*[15] Restricting once more to the corresponding reduced matrix elements one has trivially

$$(\sigma,\tau',s',s'\|\mathbf{s}\|\sigma,\tau,s,s) = (\sigma,\tau',s',s'\|\mathbf{t}\|\sigma,\tau,s,s) = \delta_{\tau'\tau}\delta_{s's}[s(s+1)(2s+1)]^{1/2}, \tag{3.5}$$

whereby we have exploited the orthonormality of the SU(2)×SU(2) basis states. Furthermore, Kemmer *et al.* have proved that[15]

$$\begin{aligned}
&(\sigma,\tau',s',s'\|T^{(1/2,1/2)}\|\sigma,\tau,s,s)\\
&= (-1)^{2s'-2s}(\sigma,\tau,s,s\|T^{(1/2,1/2)}\|\sigma,\tau',s',s')\\
&= \delta_{\tau'\tau}\{\delta_{s+1/2,s'}[(\tau-2s)(\tau+2s+3)\\
&\quad\times(s+1)(2s+1)]^{1/2} - \delta_{s-1/2,s'}[(\tau-2s+1)\\
&\quad\times(\tau+2s+2)s(2s+1)]^{1/2}\}.
\end{aligned} \tag{3.6}$$

We have made it our further task to determine the matrix elements of the SO(6) generators $U^{(1/2,1/2)}_{\alpha\beta}$ $(\alpha,\beta\in\{\frac{1}{2},-\frac{1}{2}\})$ and z_0 in the same basis (3.4). At present, all that we can say is that their action upon a state of the type (3.4) will shift the SO(5) label τ by plus or minus one unit. Indeed, z_0 and $U^{(1/2,1/2)}$ together constitute the symmetric irrep $[1,0]$ of SO(5), and by taking into account the Kronecker product reduction[18]

$$[1,0]\otimes[\tau,0] = [\tau+1,0]\oplus[\tau-1,0]\oplus[\tau,1]$$

it suffices to project the symmetric part out of the rhs in order to prove our assertion.

Unfortunately, there is no simple way to calculate directly the z_0 and $U^{(1/2,1/2)}$ matrix elements. Instead, we shall introduce in the next section yet another orthonormal state basis. For future use, let us close this section by giving expressions of the SO(5) and SO(6) Casimir operators of second degree in terms of both the SO(3) and the SU(2)×SU(2) generator basis. We find

$$\begin{aligned}
C_{2,\mathrm{SO}(5)} &= -\frac{1}{10}\sum_\mu(-1)^\mu l_\mu l_{-\mu} - \sum_\mu(-1)^\mu q_\mu q_{-\mu}\\
&= T_{1/2,1/2}T_{-1/2,-1/2} - T_{1/2,-1/2}T_{-1/2,1/2}\\
&\quad + s_0 - \mathbf{s}^2 - \mathbf{t}^2,
\end{aligned} \tag{3.7}$$

$$\begin{aligned}
C_{2,\mathrm{SO}(6)} &= C_{2,\mathrm{SO}(5)} - \frac{1}{2}\sum_\mu(-1)^\mu p_\mu p_{-\mu}\\
&= T_{1/2,1/2}T_{-1/2,-1/2} - T_{1/2,-1/2}T_{-1/2,1/2}\\
&\quad - U_{1/2,1/2}U_{-1/2,-1/2} + U_{1/2,-1/2}U_{-1/2,1/2}\\
&\quad + 2s_0 - \mathbf{s}^2 - \mathbf{t}^2 - \tfrac{1}{2}z_0^2,
\end{aligned} \tag{3.8}$$

whereby for any vector $\mathbf{u}$ the notation $\mathbf{u}^2$ stands for $-2u_{+1}u_{-1} + u_0^2 - u_0$. The Casimir operators $C_{2,\mathrm{SO}(5)}$ and $C_{2,\mathrm{SO}(6)}$ are evidently also invariants of the subalgebras SO(3) and SU(2)×SU(2). Therefore they are in an orthonormal state basis represented by diagonal matrices and since their respective eigenvalues are known to take on the values $-\tau(\tau+3)/2$ and $-\sigma(\sigma+4)/2$;[17,25] one obtains as an immediate consequence the following reduced matrix elements

$$\begin{aligned}
&\langle\sigma,\tau',v',l'\|C_{2,\mathrm{SO}(5)}\|\sigma,\tau,v,l\rangle\\
&\quad= -\tfrac{1}{2}\delta_{\tau'\tau}\delta_{l'l}(2l+1)^{1/2}\tau(\tau+3)A^{\tau}_{l}(v',v),\\
&\langle\sigma,\tau'v',l'\|C_{2,\mathrm{SO}(6)}\|\sigma,\tau,v,l\rangle\\
&\quad= -\tfrac{1}{2}\delta_{\tau'\tau}\delta_{l'l}(2l+1)^{1/2}\sigma(\sigma+4)A^{\tau}_{l}(v',v),\\
&(\sigma,\tau',s',s'\|C_{2,\mathrm{SO}(5)}\|\sigma,\tau,s,s)\\
&\quad= -\tfrac{1}{2}\delta_{\tau'\tau}\delta_{s's}\tau(\tau+3)(2s+1),\\
&(\sigma,\tau',s',s'\|C_{2,\mathrm{SO}(6)}\|\sigma,\tau,s,s)\\
&\quad= -\tfrac{1}{2}\delta_{\tau'\tau}\delta_{s's}\sigma(\sigma+4)(2s+1).
\end{aligned} \tag{3.9}$$

IV. MATRIX ELEMENTS IN AN SU(2)×SU(2)×U(1) BASIS

The operator z_0, which is a scalar under SU(2)×SU(2) [and also the zero component of the SO(3) tensor p of rank 2], thus commutes with all the SU(2) generators s_μ and t_μ ($\mu = -1,0,+1$). Hence z_0 can be considered to generate a U(1) algebra and the set $\{s_\mu,t_\mu,z_0|\mu=-1,0,1\}$ is a generator basis of an SU(2)×SU(2)×U(1) subalgebra of SO(6), which is clearly not a subalgebra of SO(5). Since the eight remaining SO(6) generators $T^{(1/2,1/2)}_{\alpha\beta}$, $U^{(1/2,1/2)}_{\alpha\beta}$ ($\alpha,\beta\in\{\frac{1}{2},-\frac{1}{2}\}$) behave as the components of two analogous bispinors under SU(2)×SU(2), they can be linearly combined to form U(1) representations. Indeed, defining

$$\begin{aligned}
A^{(1/2,1/2)}_{\alpha\beta} &= T^{(1/2,1/2)}_{\alpha\beta} - U^{(1/2,1/2)}_{\alpha\beta},\\
B^{(1/2,1/2)}_{\alpha\beta} &= T^{(1/2,1/2)}_{\alpha\beta} + U^{(1/2,1/2)}_{\alpha\beta}\\
&(\forall\alpha,\beta\in\{\tfrac{1}{2},-\tfrac{1}{2}\}),
\end{aligned} \tag{4.1}$$

we obtain by means of (3.2) that

$$\begin{aligned}
&[z_0,A^{(1/2,1/2)}_{\alpha\beta}] = A^{(1/2,1/2)}_{\alpha\beta},\\
&[z_0,B^{(1/2,1/2)}_{\alpha\beta}] = -B^{(1/2,1/2)}_{\alpha\beta}\quad(\forall\alpha,\beta\in\{\tfrac{1}{2},-\tfrac{1}{2}\}),
\end{aligned} \tag{4.2}$$

which shows that each of the components of the newly defined $A^{(1/2,1/2)}$, $B^{(1/2,1/2)}$ SU(2)×SU(2) bispinors behave as U(1) ladder operators when acting upon eigenstates of the U(1) generator z_0.

Moreover, since the five operators z_0, $\mathbf{s}^2$, s_0, $\mathbf{t}^2$, and t_0 mutually commute, the SO(6) states can be unambiguously labeled by their respective eigenvalues, in other words we can establish an SO(6) state basis of orthonormalized mutual eigenstates of these five operators. Let us denote such orthonormal states by

$$\begin{aligned}
&|\sigma,z,s,m_s,t,m_t\}\\
&\quad(m_s=-s,-s+1,\dots,s;\ m_t=-t,-t+1,\dots,t),
\end{aligned} \tag{4.3}$$

where, as before, the fact that only symmetric SO(6) irreps are considered implies that $s=t$, while from (2.8) and (3.3) it can be deduced that s takes on the values $\sigma/2$, $\sigma/2-\frac{1}{2},\dots,\frac{1}{2},0$. In order to find out what is the range of the quantum number z defined in

$$z_0|\sigma,z,s,m_s,s,m_t\} = z|\sigma,z,s,m_s,s,m_t\}, \tag{4.4}$$

let us express the SO(6) Casimir in terms of the new generator basis. Again dropping superscripts, we obtain by combining (4.1) and (3.8), and by using certain commutators between A's and B's which are derived from (3.2), the following equivalent expressions:

$$\begin{aligned}
C_{2,\mathrm{SO}(6)} &= \tfrac{1}{2}(B_{1/2,1/2}A_{-1/2,-1/2} + B_{-1/2,-1/2}A_{1/2,1/2}\\
&\quad - B_{1/2,-1/2}A_{-1/2,1/2} - B_{-1/2,1/2}A_{1/2,-1/2})\\
&\quad - \mathbf{s}^2 - \mathbf{t}^2 - \tfrac{1}{2}z_0^2 - 2z_0,
\end{aligned} \tag{4.5}$$

$$\begin{aligned}
C_{2,\mathrm{SO}(6)} &= \tfrac{1}{2}(A_{1/2,1/2}B_{-1/2,-1/2} + A_{-1/2,-1/2}B_{1/2,1/2}\\
&\quad - A_{1/2,-1/2}B_{-1/2,1/2} - A_{-1/2,1/2}B_{1/2,-1/2})\\
&\quad - \mathbf{s}^2 - \mathbf{t}^2 - \tfrac{1}{2}z_0^2 + 2z_0,
\end{aligned} \tag{4.6}$$

wherby either the operators A which shift z with $+1$, or the operators B which shift z with -1 all stand to the right. Hence denoting by $z_{\max}$ the absolute maximum that z can take in the SO(6) irrep $[\sigma,0,0]$, we learn from the action of $C_{2,\mathrm{SO}(6)}$ in the form (4.5) upon a state (4.3) with $z=z_{\max}$ the relationship $-\sigma(\sigma+4)/2 = -2s(s+1) - z_{\max}(z_{\max}+4)/2$, showing that the absolute maximum takes on the value σ and is reached when $s=0$. Similarly, the absolute minimum is $z_{\min}=-\sigma$, again on condition that $s=t=0$. By means of standard reduction techniques based upon character formulas,[26] it is possible to deduce the complete reduction rule of symmetric SO(6) irreps $[\sigma,0,0]$ into SU(2)×SU(2)×U(1) irreps. This rule can be given in two complementary forms, i.e.,

$$\begin{aligned}
&z = -\sigma,-\sigma+1,,-\sigma+2,\dots,\sigma-1,\sigma,\\
&s=t=(\sigma-|z|)/2,(\sigma-|z|)/2-1,\\
&\qquad(\sigma-|z|)/2-2,\dots,\tfrac{1}{2}\ \text{or}\ 0,
\end{aligned} \tag{4.7}$$

or

$$\begin{aligned}
&s=t=\sigma/2,(\sigma-1)/2,(\sigma-2)/2,\dots,\tfrac{1}{2},0,\\
&z=-(\sigma-2s),-(\sigma-2s)+2,\\
&\quad-(\sigma-2s)+4,\dots,(\sigma-2s)-2,(\sigma-2s).
\end{aligned} \tag{4.8}$$

We now want to calculate in the state basis (4.3) the matrix elements of all the SO(6) generators. Certain reduced matrix elements are immediately established, namely,

$$\begin{aligned}
&\{\sigma,z',s',s'\|z_0\|\sigma,z,s,s\} = \delta_{z'z}\delta_{s's}z(2s+1),\\
&\{\sigma,z',s',s'\|\mathbf{s}\|\sigma,z,s,s\} = \{\sigma,z',s',s'\|\mathbf{t}\|\sigma,z,s,s\}\\
&\qquad = \delta_{z'z}\delta_{s's}(2s+1)[s(s+1)]^{1/2}.
\end{aligned} \tag{4.9}$$

There remains to be calculated the $A^{(1/2,1/2)}$ and $B^{(1/2,1/2)}$ reduced matrix elements. This proceeds as follows.

Let $X^{(k)}$ and $Y^{(k)}$ denote two SO(3) tensors of rank k. Then[27]

$$(X^{(k)}Y^{(k)})^{(a)}_\mu = \sum_{\lambda_1,\lambda_2}\langle k\ \lambda_1\ k\ \lambda_2|a\ \mu\rangle X^{(k)}_{\lambda_1}Y^{(k)}_{\lambda_2}, \tag{4.10}$$

and

$$\begin{aligned}
[X^{(k)},Y^{(k)}]^{(a)}_\mu &= \sum_{\lambda_1,\lambda_2}\langle k\ \lambda_1\ k\ \lambda_2|a\ \mu\rangle[X^{(k)}_{\lambda_1},Y^{(k)}_{\lambda_2}]\\
&= (X^{(k)}Y^{(k)})^{(a)}_\mu - (-1)^a(Y^{(k)}X^{(k)})^{(a)}_\mu.
\end{aligned} \tag{4.11}$$

Hence when $X^{(k)}$ and $Y^{(k)}$ belong to a Lie algebra, the lhs of (4.11) can be reexpressed linearly in terms of generators, and it suffices to have a formula giving the reduced matrix of a coupled product of the type (4.10) in terms of the reduced matrix elements of the factors $X^{(k)}$ and $Y^{(k)}$, in order to obtain a nonlinear relationship between reduced matrix elements of Lie algebra generators. Such a formula is at our disposal in the form[28,29]

$$\langle l'\|(X^{(k)}Y^{(k)})^{(a)}\|l\rangle = (-1)^{l'+l+a}(2a+1)^{1/2} \times\sum_{l''}\begin{Bmatrix} k & k & a\\ l & l' & l''\end{Bmatrix}\langle l'\|X^{(k)}\|l''\rangle\langle l''\|Y^{(k)}\|l\rangle. \tag{4.12}$$

Let us next apply these formulas for $X=Y=A^{(1/2,1/2)}$ after having carried out the necessary but trivial extensions for a bitensor. Then, since $[A^{(1/2,1/2)},A^{(1/2,1/2)}]^{(a,b)}_{\mu\nu}=0$, we obtain on account of (4.11) nontrivial results only when $(a,b)=(1,0)$ or $(a,b)=(0,1)$. Substituting in (4.12) the explicit expressions for the occurring 6j symbols[27] we find

$$\begin{aligned}&-s\{\sigma,z+2,s,s\|A\|\sigma,z+1,s+\tfrac12,s+\tfrac12\}\\&\quad\times\{\sigma,z+1,s+\tfrac12,s+\tfrac12\|A\|\sigma,z,s,s\}\\&\quad+(s+1)\{\sigma,z+2,s,s\|A\|\sigma,z+1,s-\tfrac12,s-\tfrac12\}\\&\quad\times\{\sigma,z+1s-\tfrac12,s-\tfrac12\|A\|\sigma,z,s,s\}=0.\end{aligned}\tag{4.13}$$

Similarly, we have the relation

$$\begin{aligned}&-s\{\sigma,z-2,s,s\|B\|\sigma,z-1,s+\tfrac12,s+\tfrac12\}\\&\quad\times\{\sigma,z-1,s+\tfrac12,s+\tfrac12,\|B\|\sigma,z,s,s\}\\&\quad+(s+1)\{\sigma,z-2,s,s\|B\|\sigma,z-1,s-\tfrac12,s-\tfrac12\}\\&\quad\times\{\sigma,z-1,s-\tfrac12,s-\tfrac12\|B\|\sigma,z,s,s\}=0.\end{aligned}\tag{4.14}$$

As an introduction to the derivation of other relationships, attention should be drawn upon the fact that from the properties associated to Hermitian conjugation of SO(3) tensors, namely, $l^\dagger_\mu=(-1)^\mu l_{-\mu}$, $p^\dagger_\mu=(-1)^\mu p_{-\mu}$, and $q^\dagger_\mu=(-1)^\mu q_{-\mu}$ it follows with the help of (3.1) that

$$\begin{aligned}&T^\dagger_{\alpha\beta}=(-1)^{\alpha+\beta}T_{-\alpha,-\beta},\\&U^\dagger_{\alpha\beta}=(-1)^{\alpha+\beta+1}U_{-\alpha,-\beta},\\&z_0^\dagger=z_0\quad(\alpha,\beta\in\{\tfrac12,-\tfrac12\})\end{aligned}\tag{4.15}$$

and consequently, by means of (4.1) also that

$$A^\dagger_{\alpha\beta}=(-1)^{\alpha+\beta}B_{-\alpha,-\beta}\quad(\alpha,\beta\in\{\tfrac12,-\tfrac12\}).\tag{4.16}$$

The latter property may be carried over into a relationship between reduced matrix elements, i.e.,

$$\{\sigma,z-1,s',s'\|B\|\sigma,z,s,s\}^* = -\{\sigma,z,s,s\|A\|\sigma,z-1,s',s'\}.\tag{4.17}$$

Next, we shall apply the formulas (4.10)–(4.12) with $X=B^{(1/2,1/2)}$ and $Y=A^{(1/2,1/2)}$. We obtain in particular on account of the extended version of (4.11) and with the help of the commutators amongst $A^{1/2,1/2}_{\alpha\beta}$ and $B^{1/2,1/2}_{\gamma\delta}$ which are derived from (3.2), that

$$\begin{aligned}&[A^{(1/2,1/2)},B^{(1/2,1/2)}]^{(0,0)}_{00}=-4z_0,\\&[A^{(1/2,1/2)},B^{(1/2,1/2)}]^{(1,0)}_{\mu0}=-4s_\mu,\\&[A^{(1/2,1/2)},B^{(1/2,1/2)}]^{(0,1)}_{0\mu}=-4t_\mu,\\&[A^{(1/2,1/2)},B^{(1/2,1/2)}]^{(1,1)}_{\mu\nu}=0.\end{aligned}\tag{4.18}$$

Introducing finally the shorthand notations

$$\begin{aligned}&\{\sigma,z-1,s+\tfrac12,s+\tfrac12\|B\|\sigma,z,s,s\}=F(z,s),\\&\{\sigma,z-1,s-\tfrac12,s-\tfrac12\|B\|\sigma,z,s,s\}=G(z,s),\end{aligned}\tag{4.19}$$

the relations (4.13) and (4.14) are on account of (4.17) transformed into

$$-sG(z-1,s+\tfrac12)F(z,s) + (s+1)F(z-1,s-\tfrac12)G(z,s)=0\tag{4.20}$$

and its complex conjugate. In the same notation, it is straightforward to establish with the help of (4.18) and (4.10)–(4.12) three other independent relations, namely,

$$-|G(z+1,s+\tfrac12)|^2-|F(z+1,s-\tfrac12)|^2+|G(z,s)|^2 + |F(z,s)|^2=8z(2s+1)^2,\tag{4.21}$$

$$\begin{aligned}&-s|G(z+1,s+\tfrac12)|^2+(s+1)|F(z+1,s-\tfrac12)|^2\\&\quad+(s+1)|G(z,s)|^2-s|F(z,s)|^2\\&\quad=8s(s+1)(2s+1)^2,\end{aligned}\tag{4.22}$$

$$\begin{aligned}&s^2[-|G(z+1,s+\tfrac12)|^2+|F(z,s)|^2]\\&\quad+(s+1)^2[|G(z,s)|^2-|F(z+1,s-\tfrac12)|^2]=0.\end{aligned}\tag{4.23}$$

The relations (4.20)–(4.23) should be treated as recursion relations subject to certain boundary conditions. The latter follow from the reduction rules (4.7) or (4.8) combined with the definitions (4.19) and manifest themselves in the vanishing of a restricted number of F and G functions for some particular z and s values. After straightforward calculations, these functions turn out to be given (upon a possibly complex phase factor which is irrelevant to our discussion) by

$$\begin{aligned}F(z,s)&=[(s+1)(2s+1)(\sigma-z+2s+4)\\&\qquad\times(\sigma+z-2s)]^{1/2},\\G(z,s)&=[s(2s+1)(\sigma-z-2s+2)\\&\qquad\times(\sigma+z+2s+2)]^{1/2}.\end{aligned}\tag{4.24}$$

This completes after simple combination of (4.17), (4.19), and (4.24), the calculation of the reduced matrix elements of the A and B generators in the SU(2)×SU(2)×U(1) labeled SO(6) state basis. We shall now return to the SU(2)×SU(2) basis by means of explicit diagonalization of the SO(5) Casimir in the SU(2)×SU(2)×U(1) basis (4.3).

V. MATRIX ELEMENTS IN THE SU(2)×SU(2) BASIS

It is our aim to obtain in the present section the reduced matrix elements of the SO(6) generators z_0 and $U^{(1/2,1/2)}_{\alpha\beta}$ $(\alpha,\beta\in\{\tfrac12,-\tfrac12\})$ in the state basis (3.4) which makes apparent the subalgebra SO(5) and the subalgebra SU(2)×SU(2). As a start let us notice that whereas z_0 is diagonal in the orthonormal basis (4.3), it is the Casimir $C_{2,\mathrm{SO}(5)}$ which is diagonal in the orthonormal basis (3.4). The SU(2)×SU(2) contents of both bases is identically expressed by means of the labels s, m_s, and m_t. Hence diagonalizing the $C_{2,\mathrm{SO}(5)}$ Casimir in the basis (4.3) will enable us to establish explicitly the orthogonal transformation that carries that basis into the basis (3.4). It is a simple matter to verify that $C_{2,\mathrm{SO}(5)}$, defined in (3.7), can be rewritten as

$$C_{2,\mathrm{SO}(5)} = (T^{(1/2,1/2)}T^{(1/2,1/2)})^{(0,0)}_{00} - \mathbf{s}^2 - \mathbf{t}^2, \qquad (5.1)$$

hence, on account of (4.1), can also be rewritten as

$$C_{2,\mathrm{SO}(5)} = [(AA)^{(0,0)}_{00} + (BB)^{(0,0)}_{00} + (AB)^{(0,0)}_{00} + (B,A)^{(0,0)}_{00}]/4 - \mathbf{s}^2 - \mathbf{t}^2, \qquad (5.2)$$

where we have dropped superscripts again. It is now a question of applying four times the extended version of (4.12) on the rhs of (5.2), and then substituting the reduced matrix elements obtained from (4.24), to establish already the nondiagonal reduced matrix elements of the Casimir, i.e.,

$$\begin{aligned}\{\sigma,z+2,s',s'\|C_{2,\mathrm{SO}(5)}\|\sigma,z,s,s\} &= \{\sigma,z,s,s\|C_{2,\mathrm{SO}(5)}\|\sigma,z+2,s',s'\}\\ &= \delta_{s's}(2s+1)[(\sigma-z-2s)(\sigma+z+2s+4)\\ &\quad\times(\sigma-z+2s+2)(\sigma+z-2s+2)]^{1/2}/8. \qquad (5.3)\end{aligned}$$

Similarly, taking also into account that

$$\{\sigma,z,s',s'\|\mathbf{s}^2\|\sigma,z,s,s\} = \{\sigma,z,s',s'\|\mathbf{t}^2\|\sigma,z,s,s\} = \delta_{s's}s(s+1)(2s+1),$$

we can compute the diagonal matrix elements too, i.e.,

$$\{\sigma,z,s',s'\|C_{2,\mathrm{SO}(5)}\|\sigma,z,s,s\} = -\delta_{s's}(2s+1)[\sigma^2+4\sigma+4s(s+1)-z^2]/4. \qquad (5.4)$$

By setting

$$V^s_{z+k,z} = \{\sigma,z+k,s,s\|C_{2,\mathrm{SO}(5)}\|\sigma,z,s,s\} \quad (k\in\{-2,0,2\}), \qquad (5.5)$$

one should take notice of the symmetry properties

$$\begin{aligned}V^s_{z-2,z} &= V^s_{z,z-2} = V^s_{-z+2,-z} = V^s_{-z,-z+2},\\ V^s_{z,z} &= V^s_{-z,-z},\end{aligned} \qquad (5.6)$$

which, taken into consideration with the range of z values described by (4.8), show that for any acceptable fixed s value the associated V^s matrix is bisymmetric, namely, symmetric with respect to both the first and the second diagonal. By inversion of the reduction rule (3.3) one verifies that for fixed s, the SO(5) label τ runs through the possible values $2s$, $2s+1,\dots,\sigma-1,\sigma$. Hence, the matrix V^s defined in (5.5) must have all its $\sigma-2s+1$ eigenvalues in the form $-(2s+1)\tau(\tau+3)/2$ with $\tau\in\{2s,2s+1,\dots,\sigma-1,\sigma\}$. We leave it to the reader to verify this on a few low-dimensional matrices although a direct algebraical proof has become superfluous in view of the above arguments. Instead, we are much more interested in the associated set of $\sigma-2s+1$ eigenvectors of V^s which build up the orthogonal transformation matrix W^s that diagonalizes V^s, i.e.,

$$(W^s)^T V^s W^s = -\frac{(2s+1)}{2}\begin{pmatrix}\sigma(\sigma+3) & & & \\ & (\sigma-1)(\sigma+2) & & \\ & & \ddots & \\ & & & 2s(2s+3)\end{pmatrix}, \qquad (5.7)$$

and therefore also constitutes the transformation matrix carrying the basis states $|\sigma,z,s,m_s,s,m_t\}$ into basis states of the type $|\sigma,\tau,s,m_s,s,m_t)$, namely,

$$\begin{aligned}|\sigma,\tau=\sigma-i+1,s,m_s,s,m_t) &= \sum_{j=1}^{\sigma-2s+1} W^s_{ij}|\sigma,z=\sigma-2s-2j+2,s,m_s,s,m_t\}\\ &(\forall i\in\{1,2,\dots,\sigma-2s+1\},\ \forall s\in\{0,\tfrac12,1,\dots,\sigma/2\}). \qquad (5.8)\end{aligned}$$

The inverse transformation immediately follows from the orthogonality property $(W^s)^{-1}=(W^s)^T$. Letting the operator z_0 act on both sides of (5.8), and then using the inverse formula of (5.8) we arrive at the following expression for reduced matrix elements of z_0 in the SU(2)×SU(2) state basis:

$$\begin{aligned}(\sigma,\tau=\sigma-i+1,s,s\|z_0\|\sigma,\tau=\sigma-j+1,s,s) &= (\sigma,\tau=\sigma-j+1,s,s\|z_0\|\sigma,\tau=\sigma-i+1,s,s)\\ &= (2s+1)\sum_{k=1}^{\sigma-2s+1} W^s_{ik}W^s_{jk}(\sigma-2s+2k+2)\\ &(\forall i,j\in\{1,2,\dots,\sigma-2s+1\},\ \forall s\in\{0,\tfrac12,1,\dots,\sigma/2\}). \qquad (5.9)\end{aligned}$$

The calculation of the W^s matrix elements and in particular of the sum on the rhs of (5.9) is a purely algebraic problem. A complete survey of how we arrived at the solution would require too much space and therefore falls outside the scope of the present discussion. Let us just mention that the proof has been based upon complete induction and makes use of certain symmetry properties of the W^s matrices. The final result reads

$$\begin{aligned}(\sigma,\tau',s',s'\|z_0\|\sigma,\tau,s,s) &= \delta_{s's}(2s+1)\{\delta_{\tau',\tau-1}[(\sigma-\tau+1)(\sigma+\tau+3)\\ &\quad\times(\tau-2s)(\tau+2s+2)/(2\tau+1)(2\tau+3)]^{1/2}\\ &\quad+\delta_{\tau',\tau+1}[(\sigma-\tau)(\sigma+\tau+4)(\tau-2s+1)\\ &\quad\times(\tau+2s+3)/(2\tau+3)(2\tau+5)]^{1/2}\}. \qquad (5.10)\end{aligned}$$

As expected, z_0 is diagonal in the SU(2)×SU(2) labels but shifts the SO(5) label by ± 1. Notice that the ordinary matrix elements differ from the reduced ones in (5.3) only by the factor $(2s+1)$. The computation of the (reduced) matrix elements of the bispinor $U^{(1/2,1/2)}$ in the same basis is made straightforward by exploiting the relationship [see (3.2)]

$$U^{1/2,1/2}_{\alpha\beta} = [T^{(1/2,1/2)}_{\alpha\beta}, z_0] \quad (\alpha,\beta \in \{\tfrac{1}{2}, -\tfrac{1}{2}\}), \tag{5.11}$$

together with the knowledge of the reduced matrix elements (3.6) of $T^{(1/2,1/2)}$ and those of z_0 justly derived in (5.3). It is, however, not relevant here to insist on that calculation, since further on we shall only need to analyze the action of the SO(6) generators, and in particular of z_0, $U^{(1/2,1/2)}$, upon a restricted subset of states of the type $|\sigma,\tau,s,m_s,s,m_t\rangle$.

Indeed, it has been shown by Kemmer, Pursey, and Williams[15,16] that it is sufficient to consider the so-called intrinsic SU(2) $\times$ SU(2) states which constitute a two-parametric family, namely, the states

$$\chi(\sigma,\tau,\nu) = |\sigma,\tau,s = \tau/2, m_s = \tau/2 - \nu, s = \tau/2, m_t = -\tau/2\rangle$$
$$(\tau \in \{0,1,2,\ldots,\sigma\}, \quad \nu \in \{0,1,\ldots,[\tau/3]\}). \tag{5.12}$$

The restrictions imposed by the Hill–Wheeler projection technique[30] tell us that, if we succeed in expressing the action of the SO(6) generators upon the intrinsic states (5.12) either in terms of pure intrinsic states again or in terms of the action of one or more SO(3) generators l_μ ($\mu = -1,0,1$) upon intrinsic states, then we have the guarantee that the matrix elements of the SO(6) generators in the SO(3) basis can all be established in algebraic closed form. Let us recall certain results of Williams and Pursey[16]:

$$\begin{aligned} t_{+1}\chi(\sigma,\tau,\nu) &= \tfrac{1}{2} l_{+1}\chi(\sigma,\tau,\nu), \\ -[(\tau-\nu+1)/\nu]^{1/2} T^{(1/2,1/2)}_{1/2,1/2}\chi(\sigma,\tau,\nu) &= T^{(1/2,1/2)}_{-1/2,1/2}\chi(\sigma,\tau,\nu-1) \\ &= 3^{-1/2} l_{-1}\chi(\sigma,\tau,\nu-1). \end{aligned} \tag{5.13}$$

From (5.10) it follows that

$$z_0\chi(\sigma,\tau,\nu) = [(\sigma-\tau)(\sigma+\tau+4)/(2\tau+5)]^{1/2} \times |\sigma,\tau+1,\tau/2,\tau/2-\nu,\tau/2,-\tau/2\rangle. \tag{5.14}$$

Because $s = \tau/2$ there is on the rhs of (5.14) no contribution which lowers the τ value. Also, the state on the rhs is not an intrinsic state. Let us consider therefore

$$\begin{aligned} T^{(1/2,1/2)}_{-1/2,1/2}\chi(\sigma,\tau+1,\nu) &= (-1)^{\tau+\nu}\begin{pmatrix} \tau/2 & \tfrac{1}{2} & (\tau+1)/2 \\ -\tau/2+\nu & -\tfrac{1}{2} & (\tau+1)/2-\nu \end{pmatrix} \\ &\quad\times\begin{pmatrix} \tau/2 & \tfrac{1}{2} & (\tau+1)/2 \\ \tau/2 & \tfrac{1}{2} & -(\tau+1)/2 \end{pmatrix}\langle\sigma,\tau+1,\tau/2,\tau/2\|T^{(1/2,1/2)}\|\sigma,\tau+1,(\tau+1)/2,(\tau+1)/2\rangle \\ &\quad\times|\sigma,\tau+1,\tau/2,\tau/2-\nu,\tau/2,-\tau/2\rangle \\ &= (\tau-\nu+1)^{1/2}|\sigma,\tau+1,\tau/2,\tau/2-\nu,\tau/2,-\tau/2\rangle. \end{aligned} \tag{5.15}$$

Hereby we have made use of (3.6), of the explicit expressions for the $3j$ symbols and of the well-known relationship

$$\langle lm|X^{(k)}_\kappa|l'm'\rangle = (-1)^{l-m}\begin{pmatrix} l & k & l' \\ -m & \kappa & m' \end{pmatrix}\langle l\|X^{(k)}\|l'\rangle, \tag{5.16}$$

which is valid for any SO(3) tensor $X^{(k)}$. Combining (5.13), (5.14), and (5.15) we arrive at the result

$$z_0\chi(\sigma,\tau,\nu) = [(\sigma-\tau)(\sigma+\tau+4)/3(2\tau+5)(\tau-\nu+1)]^{1/2} l_{-1}\chi(\sigma,\tau+1,\nu). \tag{5.17}$$

Next, we calculate

$$\begin{aligned} &U^{(1/2,1/2)}_{-1/2,-1/2}\chi(\sigma,\tau,\nu) \\ &= \langle\sigma,\tau+1,(\tau+1)/2,(\tau+1)/2-\nu-1,(\tau+1)/2,-(\tau+1)/2|T^{(1/2,1/2)}_{-1/2,-1/2}|\sigma,\tau+1,\tau/2,\tau/2-\nu,\tau/2,-\tau/2\rangle \\ &\quad\times\langle\sigma,\tau+1,\tau/2,\tau/2-\nu,\tau/2,-\tau/2|z_0|\sigma,\tau,\tau/2,\tau/2-\nu,\tau/2,-\tau/2\rangle\chi(\sigma,\tau+1,\nu+1) \\ &= [(\nu+1)(\sigma-\tau)(\sigma+\tau+4)/(2\tau+5)]^{1/2}\chi(\sigma,\tau+1,\nu+1). \end{aligned} \tag{5.18}$$

In the intermediate steps we have made use of (5.11), (5.16), (3.6), (5.10), the orthonormality of the states and the explicit expressions for the occurring $3j$ symbols. In an analogous way, it is straightforward to prove that

$$U^{(1/2,1/2)}_{1/2,-1/2}\chi(\sigma,\tau,\nu) = [(\tau+1-\nu)(\sigma-\tau)(\sigma+\tau+4)/(2\tau+5)]^{1/2}\chi(\sigma,\tau+1,\nu). \tag{5.19}$$

The determination of the action of the two remaining $U^{(1/2,1/2)}$ components upon an intrinsic state $\chi(\sigma,\tau,\nu)$ is much more complicated than the previous cases. As an example, we shall treat the $(-1/2,1/2)$ component in more detail. In particular, the action of $U^{(1/2,1/2)}_{-1/2,1/2}$ upon $\chi(\sigma,\tau,\nu)$ produces three different contributions, i.e.,

$$\begin{aligned} U^{(1/2,1/2)}_{-1/2,1/2}\chi(\sigma,\tau,\nu) &= a|\sigma,\tau+1,(\tau+1)/2,(\tau+1)/2-\nu-1,(\tau+1)/2,-(\tau+1)/2+1\rangle \\ &\quad+ b\,|\sigma,\tau+1,(\tau-1)/2,(\tau-1)/2-\nu,(\tau-1)/2,-(\tau-1)/2\rangle + c\chi(\sigma,\tau-1,\nu). \end{aligned} \tag{5.20}$$

The calculation of the coefficients a, b, and c is straightforward and is based again upon (5.11), (5.16), (3.6), and (5.10). We obtain

$$\begin{aligned} a &= (\sigma,\tau+1,(\tau+1)/2,(\tau+1)/2-\nu-1,(\tau+1)/2,-(\tau+1)/2+1|U^{(1/2,1/2)}_{-1/2,1/2}|\sigma,\tau,\tau/2,\tau/2-\nu,\tau/2,-\tau/2) \\ &= [(\nu+1)(\sigma-\tau)(\sigma+\tau+4)/(\tau+1)(2\tau+5)]^{1/2}, \\ b &= [(\tau-\nu)(\sigma-\tau)(\sigma+\tau+4)/(\tau+1)(2\tau+3)(2\tau+5)]^{1/2}, \\ c &= -[(\tau-\nu)(\sigma-\tau+1)(\sigma+\tau+3)/(2\tau+3)]^{1/2}. \end{aligned} \tag{5.21}$$

Now, the first nonintrinsic state on the rhs of (5.20) can be easily related to an intrinsic state by considering

$$t_{+1}\chi(\sigma,\tau+1,\nu+1) = -[(\tau+1)/2]^{1/2}|\sigma,\tau+1,(\tau+1)/2,(\tau+1)/2-\nu-1,(\tau+1)/2,-(\tau+1)/2+1),$$

hence on account of (5.13)

$$|\sigma,\tau+1,(\tau+1)/2,(\tau+1)/2-\nu-1,(\tau+1)/2,-(\tau+1)/2+1) = -[2(\tau+1)]^{-1/2}l_{+1}\chi(\sigma,\tau+1,\nu+1). \tag{5.22}$$

Finally, there remains to investigate the state $|\sigma,\tau+1,(\tau-1)/2,(\tau-1)/2-\nu,(\tau-1)/2,-(\tau-1)/2)$. To that aim we calculate the action of z_0^2 upon $\chi(\sigma,\tau-1,\nu)$ in two different ways. The first way is

$$\begin{aligned} z_0^2\chi(\sigma,\tau-1,\nu) &= [(\sigma-\tau+1)(\sigma+\tau+3)/3(2\tau+3)(\tau-\nu)]^{1/2}z_0 l_{-1}\chi(\sigma,\tau,\nu) \\ &= [(\sigma-\tau+1)(\sigma+\tau+3)/3(2\tau+3)(\tau-\nu)]^{1/2}[l_{-1}z_0-\sqrt{3}\,U^{(1/2,1/2)}_{-1/2,1/2}]\chi(\sigma,\tau,\nu) \\ &= [(\sigma-\tau+1)(\sigma-\tau)(\sigma+\tau+3)(\sigma+\tau+4)/9(2\tau+3)(2\tau+5)(\tau-\nu)(\tau-\nu+1)]^{1/2} \\ &\quad\times l^2_{-1}\chi(\sigma,\tau+1,\nu) - [(\sigma-\tau+1)(\sigma+\tau+3)/(2\tau+3)(\tau-\nu)]^{1/2}U^{(1/2,1/2)}_{-1/2,1/2}\chi(\sigma,\tau,\nu), \end{aligned} \tag{5.23}$$

where we have used the commutator $[z_0,l_{-1}] = [p_0,l_{-1}] = -3^{1/2}p_{-1} = -3^{1/2}\,U^{(1/2,1/2)}_{-1/2,1/2}$, and where we applied (5.17) twice. The second way is

$$\begin{aligned} z_0^2\chi(\sigma,\tau-1,\nu) &= [(\sigma-\tau+1)(\sigma+\tau+3)/(2\tau+3)]^{1/2}z_0|\sigma,\tau,(\tau-1)/2,(\tau-1)/2-\nu,(\tau-1)/2,-(\tau-1)/2) \\ &= [(\sigma-\tau+1)(\sigma+\tau+3)/(2\tau+3)]^{1/2}\{[2(\sigma-\tau)(2\tau+2)(\sigma+\tau+4)/(2\tau+3)(2\tau+5)]^{1/2} \\ &\quad\times|\sigma,\tau+1,(\tau-1)/2,(\tau-1)/2-\nu,(\tau-1)/2,-(\tau-1)/2) \\ &\quad+[(\sigma-\tau+1)(\sigma+\tau+3)/(2\tau+3)]^{1/2}\chi(\sigma,\tau-1,\nu)\}, \end{aligned} \tag{5.24}$$

where now we have twice made use of (5.10). From (5.23) and (5.24) we obtain

$$\begin{aligned} &|\sigma,\tau+1,(\tau-1)/2,(\tau-1)/2-\nu,(\tau-1)/2,-(\tau-1)/2) \\ &\quad= [(2\tau+3)/36(\tau+1)(\tau-\nu)(\tau-\nu+1)]^{1/2}l^2_{-1}\chi(\sigma,\tau+1,\nu) \\ &\qquad-[(2\tau+3)(2\tau+5)/4(\tau+1)(\sigma-\tau)(\sigma+\tau+4)(\tau-\nu)]^{1/2}U^{(1/2,1/2)}_{-1/2,1/2}\chi(\sigma,\tau,\nu) \\ &\qquad-[(2\tau+5)(\sigma-\tau+1)(\sigma+\tau+3)/4(\tau+1)(\sigma-\tau)(\sigma+\tau+4)]^{1/2}\chi(\sigma,\tau-1,\nu). \end{aligned} \tag{5.25}$$

Finally, we substitute (5.21), (5.22), and (5.25) into (5.20) and solve the equation with respect to $U^{(1/2,1/2)}_{-1/2,1/2}\chi(\sigma,\tau,\nu)$. The result is

$$\begin{aligned} U^{(1/2,1/2)}_{-1/2,1/2}\chi(\sigma,\tau,\nu) &= -[(\tau-\nu)(\sigma-\tau+1)(\sigma+\tau+3)/(2\tau+3)]^{1/2}\chi(\sigma,\tau-1,\nu) \\ &\quad+[(\sigma-\tau)(\sigma+\tau+4)/(2\tau+5)(\tau-\nu+1)]^{1/2}/3(2\tau+3)l^2_{-1}\chi(\sigma,\tau+1,\nu) \\ &\quad-[2(\nu+1)(\sigma-\tau)(\sigma+\tau+4)/(2\tau+5)]^{1/2}/(2\tau+3)l_{+1}\chi(\sigma,\tau+1,\nu+1). \end{aligned} \tag{5.26}$$

In an analogous way one can finally prove that

$$\begin{aligned} &U^{(1/2,1/2)}_{1/2,1/2}\chi(\sigma,\tau,\nu) \\ &\quad= [\nu(\sigma-\tau+1)(\sigma+\tau+3)/(2\tau+3)]^{1/2}\chi(\sigma,\tau-1,\nu-1) \\ &\qquad-[\nu(\sigma-\tau)(\sigma+\tau+4)/(2\tau+5)(\tau-\nu+1)(\tau-\nu+2)]^{1/2}/3(2\tau+3)l^2_{-1}\chi(\sigma,\tau+1,\nu-1) \\ &\qquad-[(\sigma-\tau)(\sigma+\tau+4)/2(\tau-\nu+1)(2\tau+5)]^{1/2}(2\tau-2\nu+3)/(2\tau+3)l_{+1}\chi(\sigma,\tau+1,\nu). \end{aligned} \tag{5.27}$$

At this point we have at our disposal all the intermediate results required to apply the Hill–Wheeler projection[30] onto the physical basis.

VI. SO(6) GENERATOR REPRESENTATION IN THE PHYSICAL BASIS

We have arrived at the point whereby we can complement the results of Williams and Pursey[16] for SO(5) generators, with the explicit matrix elements in the physical basis (2.11) of the SO(6) generators p_μ ($\mu = -2, -1,...,2$) which are not contained in SO(5). In fact, the physical states are projected out of the intrinsic states by means of an invariant integral over the group manifold of SO(3), i.e.,

$$|\sigma,\tau,\nu,l,m\rangle = \int D^{l^*}_{m,K}(\Omega)\chi_\Omega(\sigma,\tau,\nu)d\Omega, \qquad (6.1)$$

whereby $D^{l}_{m,K}(\Omega)$ is an ordinary rotation matrix and $\chi_\Omega(\sigma,\tau,\nu)$ is the intrinsic SU(2)×SU(2) state on which the rotation Ω is applied. From (6.1)

$$\chi(\sigma,\tau,\nu) = \sum_l (2l+1)|\sigma,\tau,\nu,l,m\rangle \qquad (6.2)$$

follows as a trivial corollary. Now the action of a particular p component, say p_0, can be brought into the following form:

$$\begin{aligned} p_0|\sigma,\tau,\nu,l,m\rangle &= \int D^{l^*}_{mK}(\Omega)p_0\chi_\Omega(\sigma,\tau,\nu)d\Omega \\ &= \sum_{\mu,l'}\langle l\, m\, 2\, 0|l'\, m\rangle\langle l\, K\, 2\mu|l'\, K+\mu\rangle \\ &\quad\times\int D^{l'^*}_{m,K+\mu}(\Omega)p_\mu(\Omega)\chi_\Omega(\sigma,\tau,\nu)d\Omega \\ &\quad (K=\tau-3\nu). \end{aligned} \qquad (6.3)$$

But actually, we know how to express the action of all the $p_\mu(\Omega)$ upon $\chi_\Omega(\sigma,\tau,\nu)$ [or, what amounts to the same thing, the effect of all p_μ upon $\chi(\sigma,\tau,\nu)$], in terms of intrinsic states or of angular momentum operators acting upon intrinsic states. By substituting into the rhs of (6.3) the intermediate results (5.17), (5.18), (5.19), (5.26), and (5.27), after having made the changes of notation according to (3.1), what remains to be done is the calculation of integrals of the type

$$\mathcal{T}^{(a)}_{\pm} = \int D^{l'^*}_{m,K+\mu}(\Omega)(l_{\pm 1})^a\chi_\Omega(\sigma,\tau,\nu)d\Omega \qquad (a\in\{1,2\}). \qquad (6.4)$$

With the help of (6.2) we obtain in a straightforward manner:

$$\begin{aligned} \mathcal{T}^{(1)}_{\pm} &= [l'(l'+1)]^{1/2}\langle l'\, K+\mu\mp 1\, 1\pm 1|l'\, K+\mu\rangle|\sigma,\tau,\nu,l',m\rangle, \\ \mathcal{T}^{(2)}_{\pm} &= l'(l'+1)\langle l'\, K+\mu\mp 2\, 1\pm 1|l'\, K+\mu\mp 1\rangle\langle l'\, K+\mu\mp 1\, 1\pm 1|l'\, K+\mu\rangle|\sigma,\tau,\nu,l',m\rangle. \end{aligned} \qquad (6.5)$$

It is now readily established by means of (6.3) that

$$\begin{aligned} p_0&|\sigma,\tau,\nu,l,m\rangle \\ &= \sum_{l'}\langle l\, m\, 2\, 0|l'\, m\rangle\Big\{\langle l\, K\, 2\, 2|l' K+2\rangle\Big[[\nu(\sigma-\tau+1)(\sigma+\tau+3)/(2\tau+3)]^{1/2}|\sigma,\tau-1,\nu-1,l',m\rangle \\ &\quad-[l'(l'+1)/3(2\tau+3)][\nu(\sigma-\tau)(\sigma+\tau+4)/(2\tau+5)(\tau-\nu+1)(\tau-\nu+2)]^{1/2} \\ &\quad\times\langle l'\, K+4\, 1\, -1|l'\, K+3\rangle\langle l'\, K+3\, 1-1|l'\, K+2\rangle|\sigma,\tau+1,\nu-1,l',m\rangle \\ &\quad-[l'(l'+1)(\sigma-\tau)(\sigma+\tau+4)/2(\tau-\nu+1)(2\tau+5)]^{1/2}(2\tau-2\nu+3)/(2\tau+3) \\ &\quad\times\langle l'\, K+1\, 1\, 1|l'\, K+2\rangle|\sigma,\tau+1,\nu,l',m\rangle\Big] \\ &\quad+\langle l\, K\, 2\, 1|l'\, K+1\rangle[(\tau+1-\nu)(\sigma-\tau)(\sigma+\tau+4)/(2\tau+5)]^{1/2}|\sigma,\tau+1,\nu,l',m\rangle \\ &\quad+\langle l\, K\, 2\, 0|l'\, K\rangle[l'(l'+1)(\sigma-\tau)(\sigma+\tau+4)/3(2\tau+5)(\tau-\nu+1)]^{1/2} \\ &\quad\times\langle l'\, K+1\, 1\, -1|l'\, K\rangle|\sigma,\tau+1,\nu,l',m\rangle \\ &\quad+\langle l\, K\, 2\, -1\, K-1\rangle\Big[-[(\tau-\nu)(\sigma-\tau+1)(\sigma+\tau+3)/(2\tau+3)]^{1/2}|\sigma,\tau-1,\nu,l',m\rangle \\ &\quad+[l'(l'+1)/3(2\tau+3)][(\sigma-\tau)(\sigma+\tau+4)/(2\tau+5)(\tau-\nu+1)]^{1/2} \\ &\quad\times\langle l'\, K+1\, 1\, -1|l'\, K\rangle\langle l'\, K\, 1\, 1-1|l'\, K-1\rangle|\sigma,\tau+1,\nu,l',m\rangle \\ &\quad-[2l'(l'+1)(\nu+1)(\sigma-\tau)(\sigma-\tau+4)/(2\tau+5)]^{1/2}(2\tau+3) \\ &\quad\times\langle l'\, K-2\, 1\, 1|l'\, K-1\rangle|\sigma,\tau+1,\nu+1,l',m\rangle\Big] \\ &\quad+\langle l\, K\, 2\, -2|l'\, K-2\rangle[(\nu+1)(\sigma-\tau)(\sigma+\tau+4)/(2\tau+5)]^{1/2}|\sigma,\tau+1,\nu+1,l',m\rangle\Big\} \quad (K=\tau-3\nu). \end{aligned} \qquad (6.6)$$

Since the physical basis is nonorthonormal in all the labels, we have to take into account the overlap integrals $A^{\tau}_{l}(\nu',\nu)$ in (2.12) in order to derive a formula for the reduced p-matrix elements which is completely analogous to (2.14) and (2.15). The final result reads

$$
\begin{aligned}
&(2l'+1)^{-1/2}\langle\sigma\,\tau'\,\nu'\,l'\|p\|\sigma\,\tau\,\nu\,l\rangle \\
&\quad=\delta_{\tau',\tau+1}[(\sigma-\tau)(\sigma+\tau+4)/(2\tau+5)]^{1/2}\{-[l'(l'+1)/3(2\tau+3)] \\
&\qquad\times[\nu/(\tau-\nu+1)(\tau-\nu+2)]^{1/2}\langle l\,K\,2\,2|l'\,K+2\rangle\langle l'\,K+4\;1\;-1|l'\,K+3\rangle \\
&\qquad\times\langle l'\,K+3\;1\;-1|l'\,K+2\rangle A_{l'}^{\tau+1}(\nu',\nu-1) \\
&\qquad+(\tau-\nu+1)^{-1/2}[-[l'(l'+1)/2]^{1/2}(2\tau-2\nu+3)/(2\tau+3)\langle l\,K\,2\,2|l'\,K+2\rangle\langle l'\,K+1\;1\;1|l'\,K+2\rangle \\
&\qquad+(\tau-\nu+1)\langle l\,K\,2\,1|l'\,K+1\rangle+[l'(l'+1)/3]^{1/2}\langle l\,K\,2\,0|l'\,K\rangle\langle l'\,K+1\;1\;-1|l'\,K\rangle \\
&\qquad+l'(l'+1)/3(2\tau+3)\langle l\,K\,2\;-1|l'\,K-1\rangle\langle l'\,K+1\;1\;-1|l'\,K\rangle\langle l'\,K\,1\;-1\;l'\,K-1\rangle]A_{l'}^{\tau+1}(\nu',\nu) \\
&\qquad+(\nu+1)^{1/2}[-[2l'(l'+1)]^{1/2}/(2\tau+3)\langle l\,K\,2\;-1|l'\,K-1\rangle \\
&\qquad\times\langle l'\,K-2\;1\;1|l'\,K-1\rangle+\langle l\,K\,2\;-2|l'\,K-2\rangle]A_{l'}^{\tau+1}(\nu',\nu+1)\} \\
&\qquad+\delta_{\tau',\tau-1}[(\sigma-\tau+1)(\sigma+\tau+3)/(2\tau+3)]^{1/2}\{\nu^{1/2}\langle l\,K\,2\,2|l'\,K+2\rangle A_{l'}^{\tau-1}(\nu',\nu-1) \\
&\qquad-(\tau-\nu)^{1/2}\langle l\,K\,2\;-1|l'\,K-1\rangle A_{l'}^{\tau-1}(\nu',\nu)\} \\
&\qquad(K=\tau-3\nu;\ \nu\in\{0,1,\dots,[\tau/3]\};\ \ l\in\{2K,2K-2,2K-3,\dots,K\}),
\end{aligned}
\tag{6.7}
$$

where one possible expression of the function $A_l^\tau(\nu',\nu)$ has been given in (2.13).

VII. CONCLUSIONS

We have succeeded in establishing all the matrix elements of the generators of the SO(6) Lie algebra in the physical SO(3) basis. For those interested in extending the interacting boson model, these matrix elements are of great importance for the computation of the eigenvalues of SO(3) scalars in the enveloping algebra of SO(6). As an example there exists an SO(3) scalar of third degree in the generators of the form $((pl)^1 l)^0$ and its eigenvalues are upon a global l-dependent factor derived by diagonalizing with respect to the labels τ and ν the reduced matrix elements $\langle\sigma,\tau',\nu',l\|p\|\sigma,\tau,\nu,l\rangle$. The calculation of the spectra of other SO(3) scalars will require the nondiagonal elements with respect to the label l, and also the reduced matrix elements of the q generators which belong to the SO(5) subalgebra.

Hence, by the present work we have achieved results which permit to treat the two dynamical symmetries SU(3) and SO(6) in a similar way. There now remains to analyze the dynamical symmetry SU(5). We hope to report on this case in the near future.

[1]A. Arima and F. Iachello, Ann. Phys. (NY) **99**, 253 (1976).
[2]A. Arima and F. Iachello, Ann. Phys. (NY) **111**, 201 (1978).
[3]A. Arima and F. Iachello, Ann. Phys. (NY) **123**, 468 (1979).
[4]K. Heyde, P. van Isacker, M. Waroquier, and J. Moreau, Phys. Rev. C **29**, 1420 (1984).
[5]R. F. Casten, P. von Brentano, K. Heyde, P. van Isacker, and J. Jolie, Nucl. Phys. A **242**, 289 (1985).
[6]Hong-Zhou Sun, Mei Zhang, and Da Hsuan Feng, Phys. Lett. B **163**, 7 (1985).
[7]G. Vanden Berghe, H. DeMeyer, and P. van Isacker, Phys. Rev. C **32**, 1049 (1985).
[8]B. R. Judd, W. Miller, Jr., J. Patera, and P. Winternitz, J. Math. Phys. **15**, 1787 (1974), and references therein.
[9]H. DeMeyer, G. Vanden Berghe, and J. W. B. Hughes, J. Math. Phys. **22**, 2360, 2366 (1981).
[10]J. Van der Jeugt, H. DeMeyer, and G. Vanden Berghe, J. Math. Phys. **24**, 1025 (1983).
[11]H. DeMeyer, G. Vanden Berghe, and J. Van der Jeugt, J. Math. Phys. **26**, 3109 (1985).
[12]G. Vanden Berghe and H. DeMeyer, J. Math. Phys. **21**, 1902 (1980); **22**, 2326 (1981).
[13]H. DeMeyer and G. Vanden Berghe, J. Math. Phys. **21**, 1906, 1913 (1980).
[14]J. Van der Jeugt, J. Phys. A **19**, L463 (1986).
[15]N. Kemmer, D. L. Pursey, and S. A. Williams, J. Math. Phys. **9**, 1224 (1968).
[16]S. A. Williams and D. L. Pursey, J. Math. Phys. **9**, 1230 (1968).
[17]F. Iachello, in *Nuclear Structure*, edited by K. Abrahams, K. Allaart, and A. E. L. Dieperink (Plenum, New York, 1981), pp. 53–89.
[18]B. G. Wybourne, *Symmetry Principles and Atomic Spectroscopy* (Wiley–Interscience, New York, 1970).
[19]W. G. McKay and J. Patera, *Table of Dimensions, Indices and Branching Rules for Representations of Simple Lie Algebras* (Dekker, New York, 1981).
[20]B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill, New York, 1963).
[21]J. P. Elliott, Proc. R. Soc. London Ser. A **245**, 128 (1958).
[22]G. Vanden Berghe and H. DeMeyer, J. Math. Phys. **25**, 772 (1984).
[23]A. P. Stone, J. Math. Phys. **11**, 29 (1970).
[24]T. H. Seligman and R. T. Sharp, J. Math. Phys. **24**, 769 (1983).
[25]B. G. Wybourne, *Classical Groups for Physicists* (Wiley-Interscience, New York, 1974).
[26]D. E. Littlewood, *The Theory of Group Characters* (Oxford U. P., Oxford, 1958), 2nd ed.
[27]A. R. Edmonds. *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, 1957).
[28]U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic, New York, 1959).
[29]A. de-Shalit and I. Talmi, *Nuclear Shell Theory* (Academic, New York, 1963).
[30]D. L. Hill and J. A. Wheeler, Phys. Rev. **89**, 1102 (1953).

Theorems on the Jordan–Schwinger representations of Lie algebras

Shoon K. Kim
Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122

(Received 6 April 1987; accepted for publication 15 July 1987)

The Jordan–Schwinger representations of Lie algebras are discussed based on the mixed sets of creation and annihilation operators of boson or fermion type. When the representation is Hermitian, the Lie algebra is shown to be unitary for the fermion case and pseudounitary for the boson case. It is also shown that the representation of a Lie algebra leads to a projective representation of the Lie group corresponding to the Lie algebra.

I. INTRODUCTION

In 1935, Jordan[1,2] introduced the so-called Jordan mapping which is a mapping from a one-particle realization of the kinematic symmetry into field operators of either boson or fermion type. The characteristic property of the mapping is that it preserves the action of commutation of matrices. In 1952 Schwinger[3] introduced a highly ingenious treatment of the rotation group by representing the matrix generators $\{T\}$ in terms of their bilinear forms $\{\hat{T}\}$ with respect to boson annihilation and creation operators. Since this representation is equivalent to the Jordan mapping it is often called Jordan–Schwinger mapping or simply Schwinger representation. It has been widely used to provide a fairly effortless treatment of representations of Lie groups. For example, a great deal of work has been carried out on the mean field Hamiltonian $\hat{H}$ of a many body system in the second quantized field formalism, by regarding it as the Schwinger representation of a matrix generator H of the spectrum generating algebra (SGA).[4,5] Recently, the Schwinger representation has also been used to construct the algebraic Hamiltonian of the vibron models which describe the rotation–vibration spectra of nuclei[6] and the rotation–vibration spectra of molecules.[7–9]

In the present paper we shall discuss some of the basic properties of the Schwinger representation of a Lie algebra $\mathscr{L}^{(s)}$ of $s\times s$ matrices based on either fermion or boson operators. In spite of the difference in their commutation relations, the general properties of the two cases are very much parallel, in particular, if each set of particle operators describing the representation is pure, i.e., all members of each set are either annihilation operators or creation operators.[10] However, if the representations are based on mixed sets of annihilation and creation operators [see (2.4) or (3.2)], some of the properties (in particular the symmetry property) are affected by the difference in the commutation relations. For example, let T be a basis element of $\mathscr{L}^{(s)}$, then if the fermion representation $\hat{T}$ is Hermitian, the corresponding matrix generator T is also Hermitian, whereas if the boson representation $\hat{T}$ is Hermitian, the corresponding matrix T is pseudo-Hermitian (p-Hermitian). This means that the spectrum generating algebra (SGA) of $\hat{T}$ is unitary and compact for the former, whereas it is pseudounitary (p-unitary) and hence noncompact for the latter. It will be shown also that a Schwinger representation of an algebra $\mathscr{L}^{(s)}$ leads to a projective representation of the corresponding group $L^{(s)}$ in either case, even though there exists an important simplification for the fermion representation of the simple orthogonal group SO(s).

The outline of the content is as follows. In Sec. II, we shall first discuss the boson representations based on a mixed set of boson annihilation and creation operators [see (2.4)]. For this purpose it is necessary to introduce an annihilation-like operator set B and creationlike operator set $B^{\mp}$, which are defined to satisfy the boson-type commutation relations and yet they are not mutually Hermitian conjugate. This complication arises from the fact that a commutator of two noncommuting operators is antisymmetric with respect to their exchange. Then we shall discuss the symmetry properties of the representations, transformations of the particle operators, diagonalizations of the bilinear forms, and the projective representations of Lie groups $L^{(s)}$. In Sec. III we shall discuss the fermion representations in a parallel manner as the boson case. Here we have a simplifying feature over the boson case; i.e., the mixed sets of particle operators do satisfy the anticommutation relations as in the case of pure sets. In the final section (Sec. IV), we shall discuss a couple of well-known examples guided by the general theorems developed in the previous sections.

II. BOSON REPRESENTATIONS

Let $\{b_1,b_2,...,b_s\}$ and $\{b_1^\dagger,b_2^\dagger,...,b_s^\dagger\}$ be sets of boson annihilation and creation operators satisfying the commutation relations

$$[b_1,b_j^\dagger]=\delta_{ij},\quad [b_i,b_j]=[b_i^\dagger,b_j^\dagger]=0. \tag{2.1}$$

Let $B=\{B_1,B_2,...,B_s\}$ be a mixed set of annihilation and creation operators where $B_i=b_i$ or $b_i^\dagger$. Then $[B_i,B_i^\dagger]=1$ or -1. To avoid this complication we introduce an additional set of operators $B^{\mp}=\{B_i^{\mp}\}$ defined by

$$B^{\mp}=\theta\cdot B^{+}\quad\text{or}\quad B_i^{\mp}=\theta_i B_i^\dagger,\quad i=1,2,...,s, \tag{2.2}$$

where $\theta=\|\theta_i\delta_{ij}\|$ is a diagonal matrix with the diagonal elements,

$$\theta_i=\begin{cases}1, & \text{if } B_i=b_i,\\ -1, & \text{if } B_i=b_i^\dagger.\end{cases}$$

Since these satisfy the boson-type commutation relations,

$$[B_i,B_j^{\mp}]=\delta_{ij},\ [B_i,B_j]=[B_i^{\mp},B_j^{\mp}]=0, \tag{2.3}$$

we may call the set $B=\{B_i\}$ the annihilationlike operator set and $B^{\mp}=\{B_i^{\mp}\}$ the creationlike operator set. Let

 0022-2488/87/112540-06$02.50

$l(m)$ be the number of annihilation (creation) operators contained in $B=\{B_i\}$, then $l+m=s$. Without loss of generality, we may take $l \geqslant m$ and the modified sets of the operators may be given explicitly as follows:

$$\{B_i\}=(b_1,b_2,\ldots,b_l,b^\dagger_{l+1}b^\dagger_{l+2},\ldots,b^\dagger_s\},$$
$$\{B_i^\mp\}=\{b^\dagger_1,b^\dagger_2,\ldots,b^\dagger_l,-b_{l+1},-b_{l+2},\ldots,-b_s\}. \tag{2.4}$$

It will be shown that these sets are particularly effective in describing pseudounitary (p-unitary) algebra $u(l,m)$ or its subalgebra. With this much preparation we shall now describe the definition of the Jordan–Schwinger representation based on the boson operators and its properties in the following.

Let us denote a Lie group of s-dimensional matrices by $L^{(s)}$ and the corresponding Lie algebra by $\mathscr{L}^{(s)}$. Let $T=\|t_{ij}\|$ be a generator of $L^{(s)}$ and define its boson representation $\hat{T}$ by

$$\hat{T}=B^\mp\cdot T\cdot B=\sum B_i^\mp t_{ij}B_j, \tag{2.5}$$

then the set of the bilinear forms $\{\hat{T}\}$ for a basis $\{T\}$ of $\mathscr{L}^{(s)}$ defines the Schwinger representation of $\mathscr{L}^{(s)}$. The basic properties are the following.

(i) It provides a faithful representation of $\mathscr{L}^{(s)}$.

(ii) The basic commutation relations of $\hat{T}$ with the particle operators are

$$[\hat{T},B_i]=-\sum t_{ij}B_j, \tag{2.6a}$$

$$[\hat{T},B_i^\mp]=\sum B_j^\mp t_{ji}, \tag{2.6b}$$

where $i=1,2,\ldots,s$. The set of equations (2.6a) is very useful for finding a closed set of mixed operators $\{B_i\}$ from a given bilinear form expressed in terms of the original particle sets $\{b_i\}$ and $\{b^\dagger_i\}$. This is particularly so for a mean field Hamiltonian $\hat{H}$ in a second quantized formalism, since it is infinite dimensional and may be expressed by a sum of commuting bilinear forms of finite dimensions with use of (2.6a). It should be noted, however, that this procedure determines $\hat{H}$ up to an additive constant, since addition of a constant to $\hat{T}$ on the left-hand side of (2.6) does not affect the right-hand side.

(iii) Let us assume that each element U of the group $L^{(s)}$ can be represented by an appropriate generator T of $L^{(s)}$ in an exponential form $U=\exp[-iT]$. Then the Schwinger operator corresponding to U is defined by

$$\check{U}=\exp[-i\hat{T}]. \tag{2.7}$$

Under a similarity transformation by $\check{U}$, the particle operators transform as follows:

$$\check{U}B_i\check{U}^{-1}=\sum(U^{-1})_{ij}B_j,\quad \check{U}B_i^\mp\check{U}^{-1}=\sum B_j^\mp U_{ji}, \tag{2.8}$$

which follow from the basic commutation relations (2.6). These mean that $B=\{B_i\}$ and $B^\mp=\{B_i^\mp\}$ transform contragrediently. An immediate consequence of (2.8) is that the bilinear form corresponding to the unit matrix

$$B^\mp\cdot B=\sum b^\dagger_i\theta_i b_i-m \tag{2.9}$$

is invariant under a similarity transformation with $\check{U}$. In fact, it commutes with any bilinear form $\hat{T}$ of $L^{(s)}$ since it is a representation of the unit matrix. One of the most important bilinear forms we encounter in the representation is a number operator $\hat{N}_i=b^\dagger_i b_i$ of the ith particle. The total boson number operator $\hat{N}$ is defined by

$$\hat{N}=\sum b^\dagger_i b_i=\sum B_i^\mp\theta_i B_i-m. \tag{2.10}$$

It is not an invariant of the group unless $L^{(s)}=L^{(l)}\times L^{(m)}$, since $[\hat{T},\hat{N}]=0$ requires that $[T,\theta]=0$. In a special case of a simple orthogonal group SO(s) where $\theta=1$, we have additional invariants

$$B^\dagger\cdot B^\dagger,\ B\cdot B,$$

which are important in constructing the algebraic Hamiltonians for vibron models.[9]

(iv) Let $\hat{T}$ be a Hermitian bilinear form, then its spectrum generating algebra is a pseudounitary algebra u(l,m) (or its subalgebra) since the corresponding matrix generator T is pseudo-Hermitian, satisfying[11]

$$T^\dagger=\theta T\theta, \tag{2.11a}$$

where θ is the diagonal matrix defined in (2.2). The transformation operator $\check{U}=\exp[-i\hat{T}]$ is obviously unitary but the corresponding matrix $U=\exp[-iT]$ is pseudounitary, satisfying

$$U^\dagger\theta U=\theta. \tag{2.11b}$$

Note also that the matrix corresponding to $(\check{U})^\dagger$ is given by $U^{-1}=\theta U^\dagger\theta$.

(v) The transform of a bilinear form $\hat{H}=B^\mp\cdot H\cdot B$ of a matrix H under $\check{U}$ is given by the bilinear form of the transformed matrix $H'=UHU^{-1}$, i.e.,

$$\check{U}\hat{H}\check{U}^{-1}=B^\mp\cdot H'\cdot B. \tag{2.12}$$

This equation is essential for diagonalization of the operator $\hat{H}$ through diagonalization of its matrix H.

Let $\hat{H}$ be a Hermitian operator, then H is p-Hermitian. If $\hat{H}$ is diagonalized by a unitary operator $\check{U}=\exp(-i\hat{T})$ then H is diagonalized by a p-unitary matrix $U=\exp(-iT)$ (and vice versa). Since a p-Hermitian matrix remains so under a p-unitary transformation, a necessary condition for a p-Hermitian matrix H to be diagonalized by a p-unitary transformation is that the eigenvalues of H are all real. This follows since the diagonal elements of a p-Hermitian matrix are all real as in the case of a Hermitian matrix. The necessary condition is not always satisfied by a p-Hermitian matrix; e.g., the matrix $K_x=\left[\begin{smallmatrix}0&i\\i&0\end{smallmatrix}\right]$ is p-Hermitian but its eigenvalues are all imaginary. This means that a Hermitian operator $\hat{H}$ may not be necessarily diagonalized by a unitary operator $\check{U}$. (See example A in Sec. IV.) This is closely related to the fact that a compact generator cannot be transformed to a noncompact generator by a unitary transformation.

(vi) If there exists a transformation matrix U which diagonalizes H we have

$$\check{U}\hat{H}\check{U}^{-1}=\sum B_i^{\mp}h_iB_i$$

$$=\sum \theta_ih_i\Big(\hat{N}_i+\frac{1}{2}\Big)-\frac{1}{2}\,\mathrm{tr}\,H, \qquad (2.13)$$

where h_i is an eigenvalue of H and $\hat{N}_i=b_i^\dagger b_i$ is the number operator. It is often more convenient to introduce the canonical set of particle operators for $\hat{H}$ by

$$\breve{B}_i=\check{U}^{-1}B_i\check{U}=\sum U_{ij}B_j,$$
$$\breve{B}_i^{\mp}=\check{U}^{-1}B_i^{\mp}\check{U}=\sum B_j^{\mp}(U^{-1})_{ji}, \qquad (2.14)$$

then we have a canonical form of $\hat{H}$, equivalent to (2.13),

$$\hat{H}=\sum \breve{B}_i^{\mp}h_i\breve{B}_i=\sum \theta_ih_i\Big(\breve{N}_i+\frac{1}{2}\Big)-\frac{1}{2}\,\mathrm{tr}\,H, \qquad (2.15)$$

where $\breve{N}_i$ is the canonical number operator equivalent to $\hat{N}_i$. Note that θ_ih_i plays the role of h_i of the pure case. It should also be noted that $\breve{B}_i$ and $\theta_i\breve{B}_i^{\mp}$ are not mutually Hermitian adjoint unless $\check{U}$ is unitary and their vacuum state is defined by

$$|0\rangle'=\check{U}^{-1}|0\rangle, \qquad (2.16)$$

where $|0\rangle$ is the vacuum state for the original particle operators, $\{b_i,b_i^\dagger\}$.

(vii) The reverse operator of $\hat{T}$ is defined by reversing the order of the particle operators in $\hat{T}$,

$$\mathrm{Rev}\,\hat{T}\equiv B\cdot T\cdot B^{\mp}=\hat{T}+\mathrm{tr}\,T. \qquad (2.17)$$

The additive constant vanishes if $\hat{T}$ belongs to a pseudo-orthogonal algebra so(l,m), i.e.,

$$T^{\sim}=-\theta T\theta, \qquad (2.18)$$

where $T^{\sim}$ is the transpose of T. The concept of the reverse operator (first introduced by Dirac[12]) seems to give very little effect on the boson representation but it provides a far reaching consequence for the fermion representation of so(s) algebra as will be discussed in the next section.

(viii) The operator $\check{U}$ provides a projective representation of $U\in L^{(s)}$. Let U, V, $W\in L^{(s)}$ then

$$\check{U}\check{V}=k\,\check{W},\quad \text{if}\quad UV=W, \qquad (2.19)$$

where k is a constant called the projective factor of $\check{U}$ and $\check{V}$. The proof is simple; by repeated use of (2.8) one shows that $\check{W}^{-1}\check{U}\check{V}$ commutes with all the particle operators $\{B_i,B_i^{\mp}\}$ and hence is a constant. For a special case of a p-unitary group U(l,m) we have

$$|k|=1 \qquad (2.20)$$

since $\check{U}$, $\check{V}$, and $\check{W}$ are all unitary. The properties of the factor system for (2.20) are still under investigation.

III. FERMION REPRESENTATIONS

The representation based on bosons discussed in the previous section will be extended to the representation based on fermion operators $\{a_i, a_i^\dagger;\, i=1,2,\ldots,s\}$, which satisfy the anticommutation relations,

$$[a_i,a_j^\dagger]_+=\delta_{ij},\quad [a_i,a_j]_+=[a_i^\dagger,a_j^\dagger]_+=0. \qquad (3.1)$$

Since the anticommutators are symmetric with respect to the annihilation and creation operators, a mixed set defined by

$$\{A_i\}=\{a_1,a_2,\ldots,a_l,a_{l+1}^\dagger,\ldots,a_s^\dagger\} \qquad (3.2)$$

automatically satisfies the anticommutation relations

$$[A_i,A_j^\dagger]_+=\delta_{ij},\quad [A_i,A_j]_+=[A_i^\dagger,A_j^\dagger]_+=0. \qquad (3.3)$$

Here the set $\{A_i^\dagger\}$ plays the role of the set $\{B_i^{\mp}\}$ in the boson case and most of the results obtained in Sec. II hold for the present case if we replace $\{B_i,B_i^{\mp}\}$ by $\{A_i,A_i^\dagger\}$ (except for the symmetry properties of the representations) with trivial modifications.

Let us define the fermion representation of a matrix $T\in L^{(s)}$, analogously to the boson representation (2.5), by

$$\hat{T}=A^\dagger\cdot T\cdot A=\sum A_i^\dagger t_{ij}A_j. \qquad (3.4)$$

Then, if $\hat{T}$ is Hermitian, so is the matrix T and hence the transformation operator $\check{U}=\exp[-i\hat{T}]$ and the corresponding matrix $U=\exp[-iT]$ are both unitary. Thus the spectrum generating algebra of a Hermitian operator $\hat{H}$ is a unitary algebra u(s) and H can be always diagonalized by a unitary operator $\check{U}$ corresponding to a transformation matrix U which diagonalizes H.

The basic commutation relations corresponding to (2.6) are

$$[\hat{T},A_i]=-\sum_j t_{ij}A_j,\quad [\hat{T},A_i^\dagger]=\sum A_j^\dagger t_{ji}. \qquad (3.5)$$

Thus one can write down the transformation law of the particle operators under $\check{U}=\exp[-i\hat{T}]$ analogous to (2.8). The invariant bilinear form corresponding to (2.9) takes the form

$$A^\dagger\cdot A=\sum a_i^\dagger\theta_ia_i+m, \qquad (3.6)$$

where $\theta_i=1$ if $A_i=a_i$ and $\theta_i=-1$ if $A_i=a_i^\dagger$. The total fermion number operator is given by

$$\hat{N}=\sum a_i^\dagger a_i=\sum A_i^\dagger\theta_iA_i+m. \qquad (3.7)$$

The canonical set $\{\breve{A}_i\}$ of the particle operators for a Hermitian bilinear form $\hat{H}=A^\dagger\cdot H\cdot A$ is defined by, analogous to (2.14),

$$\breve{A}_i=\check{U}A_i\check{U}^\dagger=\sum U_{ij}A_j, \qquad (3.8)$$

where U is the unitary matrix which diagonalizes H. These may be written explicitly, as follows:

$$\breve{A}_i=\{\breve{a}_1,\breve{a}_2,\ldots,\breve{a}_l,\breve{a}_{l+1}^\dagger,\ldots,\breve{a}_s^\dagger\}, \qquad (3.9)$$

where $\breve{a}_i$ and $\breve{a}_i^\dagger$ are the annihilation and creation operators which act on the canonical vacuum state $U^\dagger|0\rangle$. Thus the bilinear form takes the form

$$\hat{H}=\sum \breve{A}_i^\dagger h_i\breve{A}_i=\sum \theta_ih_i\Big(\breve{N}_i-\frac{1}{2}\Big)+\frac{1}{2}\,\mathrm{tr}\,H, \qquad (3.10)$$

where h_i is an eigenvalue of H and $\breve{N}_i=\breve{a}_i^\dagger\breve{a}_i$ is a canonical number operator. Note that Eqs. (3.6), (3.7), and (3.10) have the same forms as those of corresponding equations (2.9), (2.10), and (2.15) of the boson case, respectively, except for the opposite signs of the additive constants. Note, in particular, that $\theta_i h_i$ plays the role of h_i of the pure case as in the boson case.

Now, in general the Schwinger operator $\check{U} = \exp[-i\hat{T}]$ provides a projective representation of $U = \exp[-iT] \in L^{(s)}$;

$$\check{U}\check{V} = k\check{W}, \quad \text{if } UV = W, \tag{3.11}$$

which may be proved analogously in the case of boson operators. The projective factor k satisfies

$$|k| = 1 \tag{3.12}$$

for a unitary group. In a further special case of SO(s), $\check{U}$ provides a double-valued representation with

$$k = \pm 1. \tag{3.13}$$

We shall prove this important result based on the concept of reverse operators introduced by Dirac[12] for the theory of spinor representation.

For any operator O defined as a function of the noncommuting operators, its reverse operator $O^{\sim}$ is obtained from O by reversing the factors in every product in O. The same symbol $^{\sim}$ is used for the reverse of an operator as for the transpose of a matrix because there is the same product law,

$$\begin{aligned}(O_1O_2)^{\sim} &= O_2^{\sim}O_1^{\sim},\\ (O^{-1})^{\sim} &= (O^{\sim})^{-1}.\end{aligned} \tag{3.14}$$

For a bilinear form $\hat{T} = A^{\dagger}\cdot T\cdot A$ we have

$$(\hat{T})^{\sim} = -\hat{T} + \operatorname{tr} T. \tag{3.15}$$

Let T be an antisymmetric matrix then the transformation matrix $U = \exp[-iT]$ belongs to SO(s) and the Schwinger operator U satisfies

$$\check{U}^{\sim}\check{U} = 1. \tag{3.16}$$

Now (3.13) follows from (3.11) and its reverse equation $\check{V}^{\sim}\check{U}^{\sim} = k\check{W}^{\sim}$ and (3.16) (Q.E.D.). The double-valued representation does not hold for the boson case, simply because the reverse operation has little effect on $\hat{T}$ for the boson case as has been mentioned in (vii) of Sec. II.

IV. EXAMPLES

To show the effectiveness of the general theorems developed for the Schwinger representation in the previous sections we shall discuss some well-known examples, the boson representation of the SU(1.1) group and the fermion representation of the SU(2) group, guided by the general theorems.

A. SU(1.1) boson model

The matrix generators of SU(1.1) may be defined by[13]

$$K_1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad K_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad K_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{4.1}$$

These matrices are p-Hermitian, satisfying

$$K_i^{\dagger} = \theta K_i \theta; \quad \theta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{4.2}$$

Following (2.4), the sets of mixed operators are chosen as follows:

$$\begin{aligned}B &= (B_1,B_2) = (b_1,b_2^{\dagger}),\\ B^{\mp} &= (B_1^{\mp},B_2^{\mp}) = (b_1^{\dagger},-b_2).\end{aligned} \tag{4.3}$$

Then the boson representation of the basis $\{K_1,K_2,K_3\}$ is given by a set of Hermitian bilinear forms $\hat{K}_i = B^{\mp}\cdot K_i\cdot B$ or explicitly by

$$\begin{aligned}\hat{K}_1 &= i(b_1^{\dagger}b_2^{\dagger} - b_2b_1),\\ \hat{K}_2 &= b_1^{\dagger}b_2^{\dagger} + b_2b_1,\\ \hat{K}_3 &= b_1^{\dagger}b_1 + b_2b_2^{\dagger}.\end{aligned} \tag{4.4}$$

The operators $\hat{K}_1$ and $\hat{K}_2$ are noncompact since the eigenvalues of K_1 and K_2 are imaginary whereas the operator $\hat{K}_3$ is compact since K_3 has real eigenvalues.

The most general Schwinger generator of SU(1.1) which is Hermitian is given by

$$\hat{M} = \xi_1\hat{K}_1 + \xi_2\hat{K}_2 + \epsilon\hat{K}_3. \tag{4.5}$$

This operator is compact or noncompact depending on the relative magnitudes of the real coefficients (ξ_1,ξ_2,ϵ). To have some physical feeling, this generator must be compared with the Bogoliubov Hamiltonian[4,14] for the superfluid boson model given by

$$\begin{aligned}\hat{H} &= \sum_{k>0}\hat{H}(k),\\ \hat{H}(k) &= s(k^2)(b_k^{\dagger}b_k + b_{-k}b_{-k}^{\dagger})\\ &\quad + t(k^2)(b_k^{\dagger}b_{-k}^{\dagger} + b_{-k}b_k),\end{aligned} \tag{4.6}$$

where $s(k^2)$ and $t(k^2)$ are scalar functions of the wave vector k. Then $\hat{H}(k)$ is a special case of $\hat{M}$ with $b_1 = b_k$, $b_2 = b_{-k}$ and $\xi_1 = 0$, $\xi_2 = t(k^2)$, $\epsilon = s(k^2)$. For this model $\hat{M}$ is compact since $|s(k^2)| > t(k^2)|$.

The matrix M corresponding to $\hat{M}$ is given by

$$\begin{aligned}M &= \xi_1K_1 + \xi_2K_2 + \epsilon K_3\\ &= \begin{pmatrix} \epsilon & \xi \\ -\xi^* & -\epsilon \end{pmatrix}; \quad \xi = \xi_2 + i\xi_1,\end{aligned} \tag{4.7}$$

which is p-Hermitian. By a Bogoliubov transformation, the matrix M can be diagonalized into the form

$$UMU^{-1} = EK_3, \quad E = (\epsilon^2 - |\xi|^2)^{1/2}. \tag{4.8}$$

Here if $|\epsilon| > |\xi|$, E is real and its sign can be chosen such that $\epsilon E > 0$, whereas if $|\epsilon| < |\xi|$, E becomes imaginary and we set $E = i|E|$. In either case, the transformation matrix U may be given by a quasirotation[15] about a unit vector $n = (\xi_2/|\xi|, -\xi_1/|\xi|, 0)$ through an "angle" $\chi = \tanh^{-1}(|\xi|/\epsilon)$,

$$U = \exp[-(i/2)\chi(n_1K_1 + n_2K_2)] \tag{4.9a}$$

or explicitly by

$$U = N\begin{pmatrix} \epsilon + E & \xi \\ \xi^* & \epsilon + E \end{pmatrix} \equiv \begin{pmatrix} u & v_1 \\ v_2 & u \end{pmatrix}, \tag{4.9b}$$

where $N^{-2} = 2E(\epsilon + E)$ and $u^2 - v_1v_2 = 1$. Note that

$$U^{-1} = \theta U^{\dagger}\theta. \tag{4.10}$$

It is also noted that the explicit form (4.9b) can be directly obtained by means of a general theory of matrix diagonalization developed by the author.[16]

When $|\epsilon| > |\xi|$, the angle χ is real so that $\check{U}$ is unitary, and hence U is p-unitary, i.e., $U^{\dagger} = \theta U^{-1}\theta$. This is expected since both $\hat{M}$ and $\hat{K}_3$ are compact. When $|\epsilon| < |\xi|$, the angle χ

becomes complex so that $\breve{U}$ cannot be unitary. This is also expected since $\hat{M}$ is noncompact this time. This is a direct consequence of the statement given in (V) of Sec. II that a p-Hermitian matrix M with complex eigenvalues cannot be diagonalized by a p-unitary matrix. It is also noted that the noncompact $\hat{M}$ can be transformed to the noncompact operator $|E|(\xi_1\hat{K}_1 + \xi_2\hat{K}_2)/|\xi|$ by a unitary operator $\breve{U}$ with $\chi = \tanh^{-1}(\epsilon/|\xi|)$ in (4.9a).

In terms of the matrix U of (4.9b) which diagonalizes M, the canonical set of the particle operators for $\hat{M}$ is given by, using (2.14),

$$\breve{B}_1 = ub_1 + v_1 b_2^\dagger, \quad \breve{B}_2 = v_2 b_1 + ub_2^\dagger, \qquad \breve{B}_1^\mp = ub_1^\dagger + v_2 b_2, \quad \breve{B}_2^\mp = -v_1 b_1^\dagger - ub_2. \tag{4.11}$$

The canonical form of $\hat{M}$ is given by

$$\hat{M} = E(\breve{B}_1^\mp \breve{B}_1 - \breve{B}_2^\mp \breve{B}_2) = E(\breve{N}_1 + \breve{N}_2 + 1), \tag{4.12}$$

which is consistent with the general expression (2.15).

B. The SU(2) fermion model

The mean-field Hamiltonian $\hat{H}$ for the Bardeen–Cooper–Schrieffer (BCS) model[5,17] is given by

$$\hat{H} = \sum_k \hat{H}(k),$$

$$\hat{H}(k) = \epsilon(k)(a_{k\uparrow}^\dagger a_{k\uparrow} - a_{-k\downarrow} a_{-k\downarrow}^\dagger) + \Delta(k) a_{k\uparrow}^\dagger a_{-k\downarrow}^\dagger + \Delta^*(k) a_{-k\downarrow} a_{k\uparrow}, \tag{4.13}$$

where $\epsilon(k)$ is one-particle energy and $\Delta(k)$ is the energy gap function of the wave vector k; $a_{k\uparrow}$ and $a_{k\uparrow}^\dagger$ are annihilation and creation operators of the electron with spin up. By applying (3.5) to the bilinear form $\hat{H}(k)$ of (4.13), one can express it in the form $\hat{H}(k) = A^\dagger \cdot H(k) \cdot A$ with

$$A = \{A_1, A_2\} = \{a_{k\uparrow}, a_{-k\downarrow}^\dagger\} \tag{4.14}$$

and

$$H(k) = \begin{pmatrix} \epsilon & \Delta \\ \Delta^* & -\epsilon \end{pmatrix} = \Delta_1\sigma_1 - \Delta_2\sigma_2 + \epsilon\sigma_3, \tag{4.15}$$

where $\Delta = \Delta_1 + i\Delta_2$, and $(\sigma_1, \sigma_2, \sigma_3)$ are the Pauli matrices and we have suppressed the k dependence of ϵ and Δ.

Since the given bilinear form $\hat{H}(k)$ is Hermitian, the matrix $H(k)$ is also Hermitian as it should be and hence $H(k)$ can be diagonalized by a unitary transformation as follows:

$$u\,H(k)U^\dagger = E\sigma_3, \quad E = (\epsilon^2 + |\Delta|^2)^{1/2}, \quad \epsilon E > 0. \tag{4.16}$$

The spectrum generating algebra (SGA) of the Hamiltonian $\hat{H}(k)$ is su(2) with the basis $(\sigma_1, \sigma_2, \sigma_3)$. The unitary matrix U which diagonalizes $H(k)$ in (4.16) is given by a rotation about a unit vector $n = (-\Delta_2/|\Delta|, -\Delta_1/|\Delta|, 0)$ through an angle $\phi = \tan^{-1}(|\Delta|/\epsilon)$,

$$U = \exp[-(i/2)\phi(n_1\sigma_1 + n_2\sigma_2)] \tag{4.17a}$$

or explicitly by[16]

$$U = N\begin{pmatrix} \epsilon + E & \Delta \\ -\Delta^* & \epsilon + E \end{pmatrix} \equiv \begin{pmatrix} u & v \\ -v^* & u \end{pmatrix}, \tag{4.17b}$$

where $N^{-2} = 2E(\epsilon + E)$ and $u^2 + |v|^2 = 1$. From (3.8) and (4.17b) the canonical set of the particle operators $\{\breve{A}_1, \breve{A}_2\} = \{\breve{a}_1, \breve{a}_2^\dagger\}$ is given by

$$\breve{a}_1 = u\,a_{k\uparrow} + va_{-k\downarrow}^\dagger, \qquad \breve{a}_2^\dagger = -v^* a_{k\uparrow} + ua_{-k\downarrow}^\dagger, \tag{4.18}$$

and the canonical form of $\hat{H}(k)$ is given by

$$\hat{H}(k) = E(\breve{a}_1^\dagger \breve{a}_1 - \breve{a}_2 \breve{a}_2^\dagger) = E(\breve{N}_1 + \breve{N}_2 - 1), \tag{4.19}$$

which is consistent with (3.10).

An easy application of (4.19) and the canonical operators of (4.18) is to calculate the thermal average of an off-diagonal element over the canonical distribution with temperature T (with energy measured from the Fermi surface)

$$\langle a_{k\uparrow} a_{-k\downarrow} \rangle = (\Delta(k)/2E(k)) \langle (1 - \breve{N}_1(k) - \breve{N}_2(k)) \rangle = (\Delta(k)/2E(k)) \tanh[E(k)/2k_B T], \tag{4.20}$$

where we have used the inverse transformation of (4.18) and restored the k dependence. As is well known, the gap function $\Delta(k)$ in the mean-field Hamiltonian is determined self-consistently from[18]

$$\Delta(k) = -\sum V_{kk'} \langle a_{k'\uparrow} a_{-k'\downarrow} \rangle, \tag{4.21}$$

where the $V_{kk'}$ are the potential parameters which are nonzero only in the vicinity of the Fermi surface.

Finally, the ground state of $\hat{H}(k)$ is given by the canonical vacuum state $\breve{U}^\dagger|0\rangle$ of $\{\breve{A}_i\}$, which is a coherent state[19] of SU(2). Using the Baker–Campbell–Hausdorff formula,[11] it is rewritten in the more familiar form[18]

$$\breve{U}^\dagger|0\rangle = \kappa(u - va_{k\uparrow}^\dagger a_{-k\downarrow}^\dagger)|0\rangle, \quad |\kappa| = 1, \tag{4.22}$$

where the coefficients u,v are defined by (4.17b) and the phase factor κ is due to the projective factor which arises since $\breve{U}$ is a projective representation of $U \in SU(2)$.

V. CONCLUDING REMARKS

In terms of mixed sets of boson annihilation and creation operators we have introduced an annihilationlike operator set $B = \{B_i\}$ and a creationlike operator set $B^\mp = \{B_i^\mp\}$ which satisfy the bosonlike commutation relations. The Schwinger representations of Lie algebra $\mathscr{L}^{(s)}$ based on these sets are discussed. It has been shown that a Hermitian bilinear form $\hat{H}$ with respect to B and $B^\mp$ belongs to a pseudounitary algebra and thus it may not be necessarily diagonalized by a unitary operator. In the case of fermion representations, a Hermitian form $\hat{H}$ always belongs to a unitary algebra and can always be diagonalized by a unitary operator $\breve{U}$.

The Schwinger operator $\breve{U} = \exp[-i\hat{T}]$ with bosons or fermions is shown to provide a projective representation of an element $U = \exp[-iT]$ of a Lie group $L^{(s)}$. It becomes a double-valued representation for SO(s) in the case of fermion operators. The properties of the factor system will be discussed in a forthcoming paper.

ACKNOWLEDGMENT

The author is grateful to Professor R. D. Levine of the Fritz Haber Research Center for Molecular Dynamics, The Hebrew University, Jerusalem, Israel for the hospitality ex-

tended to me during my stay in Israel. Part of this work has been carried out at the center.

[1]P. Jordan, Z. Phys. **94**, 531 (1935).
[2]L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics* (Addison–Wesley, London, 1981), p. 212.
[3]J. Schwinger, "On angular momentum," U.S. Atomic energy commission, NYO-3071, 1952 (unpublished). It is contained in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. Van Dam (Academic, New York, 1965).
[4]A. I. Solomon, J. Math. Phys. **12**, 390 (1971).
[5]A. I. Solomon and J. L. Birman, Phys. Lett. A **111**, 423 (1985); J. Math. Phys. **28**, 1526 (1987).
[6]A. Arima and F. Iachello, Ann. Phys. (NY) **99**, 253 (1976); **111**, 201 (1978); **123**, 468 (1979).
[7]R. D. Levine and C. G. Wulfman, Chem. Phys. Lett. **60**, 372 (1979).
[8]F. Iachello and R. D. Levine, J. Chem. Phys. **77**, 3046 (1982).
[9]S. K. Kim, I. L. Cooper, and R. D. Levine, Chem. Phys. **106**, 1 (1986).
[10]S. K. Kim, "Theorem on the Schwinger representation of Lie groups and its application to the coherent states and the vibron model," in *Proceedings of the 1st International Conference on the Phase Space* (Springer, New York, 1986), p. 314; see also "Theorem on the Schwinger representations of Lie groups and its application to the vibron model," to be published in *Proceedings of the XV International Colloquium on Group Theoretical Methods in Physics*.
[11]R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974).
[12]P. A. M. Dirac, *Spinors in Hilbert Space* (Plenum, New York, 1974).
[13]Y. Alhassid, F. Gursey, and F. Iachello, Ann. Phys. (NY) **148**, 346 (1983).
[14]N. Bogoliubov, J. Phys. (Moscow) **11**, 23 (1947).
[15]S. K. Kim, J. Math Phys. **21**, 1299 (1980).
[16]S. K. Kim, J. Math Phys. **20**, 2153 (1979).
[17]J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).
[18]R. D. Parks, *Superconductivity* (Dekker, New York, 1969), Vol. 1, p. 67.
[19]A. M. Perelomov, Sov. Phys. Usp. **20**, 703 (1977).

Canonical map approach to channeling stability in crystals. II

A. W. Sáenz
Naval Research Laboratory, Washington, D.C. 20375-5000 and Physics Department, Catholic University, Washington, D.C. 20064

(Received 17 July 1986; accepted for publication 1 April 1987)

A nonrelativistic and a relativistic classical Hamiltonian model of two degrees of freedom are considered describing the plane motion of a particle in a potential $V(x_1,x_2)$ [(x_1,x_2) = Cartesian coordinates]. Suppose $V(x_1,x_2)$ is real analytic in its arguments in a neighborhood of the line $x_2 = 0$, one-periodic in x_1 there, and such that the average value of $\partial V(x_1,0)/\partial x_2$ vanishes. It is proved that, under these conditions and provided that the particle energy E is sufficiently large, there exist for all time two distinguished solutions, one satisfying the equations of motion of the nonrelativistic model and the other those of the relativistic model, whose corresponding configuration-space orbits are one-periodic in x_1 and approach the line $x_2 = 0$ as $E \to \infty$. The main theorem is that these solutions are (future) orbitally stable at large enough E if V satisfies the above conditions, as well as natural requirements of linear and nonlinear stability. To prove their existence, one uses a well-known theorem, for which a new and simpler proof is provided, and properties of certain natural canonical maps appropriate to these respective models. It is shown that such solutions are orbitally stable by reducing the maps in question to Birkhoff canonical form and then applying a version of the Moser twist theorem. The approach used here greatly lightens the labor of deriving key estimates for the above maps, these estimates being needed to effect this reduction. The present stability theorem is physically interesting because it is the first rigorous statement on the orbital stability of certain channeling motions of fast charged particles in rigid two-dimensional lattices, within the context of models of the stated degree of generality.

I. INTRODUCTION

In a previous paper,[1] we initiated the rigorous study of the orbital stability of certain rectilinear motions occurring in the context of nontrivial classical Hamiltonian models of channeling in rigid two-dimensional crystals. In the present paper, we extend these results to a more general class of models of this type in which one cannot infer the existence of certain distinguished motions by symmetry arguments.[2] A typical example is as follows. Suppose that a fast charged particle is injected into a square array of atoms along a line parallel to the $\langle 1,1\rangle$ atomic strings and lying midway between two such strings. Under reasonable hypotheses on the repulsive interactions between the atoms of the array and the fast particle, does there exist a semi-infinite trajectory which remains close for all future times to this midline, and is the corresponding motion stable in some sense? Many numerical studies[3,4] suggest that the answers to questions of this type are affirmative. The purpose of this paper is to show rigorously that this is the case within the framework of the above generalized Hamiltonian models.

Such generalized two-dimensional models are of physical interest because they describe a rich variety of channeling phenomena occurring in three-dimensional crystals, within the domain of validity of suitable continuum-model approximations. They are of mathematical interest because it is nontrivial that such semi-infinite trajectories exist in the context of the present models, and if they do whether they are stable or unstable in some well-defined sense.

One of the two generalized models considered here is nonrelativistic (NR model) and the other relativistic (R model). They are described by the respective Hamiltonians $H_{\rm NR}$ and $H_{\rm R}$, which are smooth real-valued functions defined at each $(x_1,x_2,p_1,p_2)\in\mathbb{R}^4$ such that (x_1,x_2) is in an appropriate neighborhood of the line $x_2 = 0$ (center of a channel) in $\mathbb{R}^2$ (see Ref. 5):

$$H_{\rm NR}(x_1,x_2,p_1,p_2) = \tfrac{1}{2}(p_1^2 + p_2^2) + V(x_1,x_2)\,, \quad (1.1a)$$

$$H_{\rm R}(x_1,x_2,p_1,p_2) = \sqrt{p_1^2 + p_2^2 + 1} + V(x_1,x_2)\,. \quad (1.1b)$$

In Sec. II, we state our principal results, Theorems 1 and 2, and discuss the strategy of their proofs. Let the potential $V(x_1,x_2)$ be a real-valued function which is analytic in x_1,x_2 jointly in the above neighborhood of the line $x_2 = 0$, one-periodic in x_1 there, and such that the x_2 component $-\partial V(x_1,x_2)/\partial x_2$ of the force on the particle of interest vanishes *on the average* along this line. Theorem 1 asserts that under these conditions and provided that E is sufficiently large there exists for the NR model and the R model a distinguished solution of the respective equations of motion, whose phase-space orbit is one-periodic in x_1 and whose x_2 and p_2 coordinates tend to zero as $E \to \infty$. The last condition on V is substantially weaker than the corresponding one imposed in Ref. 1 that this force component vanish *pointwise* for $x_2 = 0$. We also remark that under the assumptions of Ref. 1 these distinguished solutions reduce to the corresponding rectilinear ones considered there. If V satisfies the conditions of Theorem 1 plus certain natural requirements of linear and nonlinear stability, Theorem 2 asserts the (future) orbital stability of the distinguished solutions in question at sufficiently large E. This is our main result.

We prove Theorems 1 and 2 in Sec. III for the NR model and in Sec. IV for the R model. We show the existence of the latter solutions by using Theorem A.1 of Appendix A, which

 0022-2488/87/112546-13$02.50

is a version of well-known mathematical results,[6,7] originally derived by averaging-theory methods. A simple proof of Theorem A.1 not appealing to such methods is given in that appendix for the sake of completeness. In the proofs of Theorems 1 and 2, we use properties of two natural canonical maps from a neighborhood of $0 \in \mathbb{R}^2$ into $\mathbb{R}^2$, one of which corresponds to the NR model and the other to the R model. At large enough E, these maps can be reduced to maps in Birkhoff normal form having the properties assumed in the version of Moser's twist theorem[8] stated as Theorem B.1 in Appendix B. The application of this version to the latter maps and elementary arguments complete the proof of Theorem 2. In order to show that the Birkhoff normal forms of the above maps have these properties, one needs estimates of certain of their derivatives at the relevant fixed points in the limit $E \to \infty$. The derivation of these estimates is considerably shorter by the power-series approach of the present paper than by the methods of Ref. 1.

Ellison and his collaborators[9] have derived important rigorous results on channeling motions by averaging-theory arguments. The theory of averaging is of very general applicability in a channeling context, but save in exceptional circumstances its conclusions are only known to hold for finite, although typically long-time intervals, as is the case for the results derived in Ref. 9. On the other hand the canonical map approach advocated in the present paper leads to stability theorems obtaining over infinite time intervals, but up to now it has only been possible to apply it to rather special channeling motions. It is natural to conjecture that much could be gained in our theoretical understanding of channeling by an adroit combination of these two methods.

II. STATEMENT OF MAIN RESULTS AND STRATEGY OF PROOF

A. Main results

We begin by stating more precisely the conditions imposed on the potential V occurring in (1.1a) and (1.1b). In the majority of our discussions, it will be assumed to possess the following two properties.

(I) $V(x_1,x_2)$ is a real-valued function which is analytic in the real variables x_1, x_2 (Ref. 10) in the closed strip

$$\mathscr{S} = \{(x_1,x_2)\in\mathbb{R}^2\colon\ |x_2|\leqslant a_0\}, \tag{2.1}$$

for some constant $a_0 > 0$.

(II) In the strip (2.1), $V(x_1,x_2)$ is periodic in x_1 with period unity and

$$A_1 = 0, \tag{2.2}$$

where

$$A_j = \int_0^1 \frac{\partial j\, V(x_1,0)}{\partial x_2^j}\, dx_1. \tag{2.3}$$

In the proof of our main result—Theorem 2—we will also assume the following.

(III) The inequalities

$$A_2 > 0, \tag{2.4a}$$

$$A_2 A_4 - \tfrac{5}{3} A_3^2 \neq 0 \tag{2.4b}$$

hold.

Properties (I) and (III) were also assumed in Ref. 1, where their motivation was discussed. In particular, Eqs. (2.4a) and (2.4b) are conditions of linear and nonlinear stability conditions, respectively. Property (II) obviously states that the x_2 component of the force on the particle of interest vanishes *on the average* along the line $x_2 = 0$.

Unless an explicit statement to the contrary is made, the symbol H in the following discussions stands for $H_{\rm NR}$ or $H_{\rm R}$.

Let V have property (I). Then at each quadruple ρ, ξ, η, E of real numbers with $(\rho,\xi)\in\mathscr{S}$ for which it exists, we define $x_i^H(t;z,\epsilon), p_i^H(t;z,\epsilon)$ $(i = 1,2)$ as a solution of the Hamiltonian equations of motion

$$\dot{x}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial x_i}, \quad i = 1,2, \tag{2.5}$$

at each t of a maximum interval in $\mathbb{R}$ containing $t = 0$, satisfying the initial conditions

$$\begin{aligned} &x_1^H(0;z,E) = \rho, \quad x_2^H(0;z,E) = \xi, \\ &p_1^H(0;z,E) > 0, \quad p_2^H(0;z,E) = \eta, \end{aligned} \tag{2.6}$$

and such that at each t in this interval $(x_1^H(t;z,E), x_2^H(t;z,E), p_1^H(t;z,E), p_2^H(t;z,E))$ lies on the energy surface $H = E$ in $\mathbb{R}^4$ and $(x_1^H(t;z,\epsilon),x_2^H(t;z,\epsilon))\in\mathscr{S}$. Here $z = (\rho,\zeta)$, where $\zeta = (\xi,\eta)$. This solution is unique in view of the smoothness properties of the right sides of these equations of motion when V has the assumed property (I).

If $V(x_1,x_2)$ has this property and in addition is one-periodic in x_1, we see by the results of the last paragraphs of Secs. III A and IV A that there exist a neighborhood $\mathscr{U} \subset \mathbb{R}^2$ of the origin [$\mathscr{U}$ is the same for $H = H_{\rm NR}$ and $H = H_{\rm R}$; see (3.7b)] and a positive constant E_0^H such that for $z\in\mathbb{R}\times\mathscr{U}$, $E > E_0^H$, system (2.5) has precisely one solution $x_i^H(t;z,E)$, $p_i^H(t;z,E)$ $(i = 1,2)$ on suitable time intervals. Moreover, by the properties of the diffeomorphisms (3.8a) and (4.7a) stated in these respective subsections, there is at each such z,E a unique time $\tau_H(z,E)$ at which this solution of (2.5), which emerges from the hyperplane $x_1 = \rho$ at $t = 0$, pierces the hyperplane $x_1 = \rho + 1$, i.e.,

$$x_1^H(\tau_H(z,E);z,E) = \rho + 1. \tag{2.7}$$

Our principal result asserts that when V has properties (I)–(III) certain distinguished solutions of system (2.5) for $H = H_{\rm NR}$, $H = H_{\rm R}$, which exist at all times and whose phase-space orbits are close to the hyperplane $x_2 = p_2 = 0$ at sufficiently high E, are orbitally stable at such E values. These solutions are defined in the next theorem. There and elsewhere in this paper, the notion of orbital stability is understood as follows.

Definition: Let $\mathscr{O}_+(z,E)$ be the phase-space orbit, if it exists, swept out during the time interval $t\geqslant 0$ by a particle of energy E subject to the initial conditions (2.6), i.e.,

$$\mathscr{O}_+(z,E) = \{(x_1(t;z,E),x_2(t;z,E),p_1(t;z,E),p_2(t;z,E)),\ 0\leqslant t < \infty\}\subset\mathbb{R}^4.$$

Suppose that the solution $x_i^H(t;z,E), p_i^H(t;z,E)$ $(i = 1,2)$ exists at some $\tilde{z} = (\tilde{\rho},\tilde{\xi})$, $\widetilde{E}$ for all $t\geqslant 0$. This solution is said to be orbitally stable if for all $\epsilon > 0$ there exists $\delta = \delta(\epsilon,\tilde{\rho},\widetilde{E})$ such that $\mathscr{O}_+(z,E)$ exists and each of its points p satisfies $d(p,\mathscr{O}_+(\tilde{z},\widetilde{E})) < \epsilon$ if $d((z,E), (\tilde{z},\widetilde{E})) < \delta$ where d is the distance function appropriate to the usual norm on $\mathbb{R}^4$ (see Ref. 5).

Theorem 1: Let V obey conditions (I), (II), and

$A_2 \neq 0$.

Then there exist a positive constant E_H and a neighborhood $\mathscr{D}_H \subset \mathscr{U}$ of $0 \in \mathbb{R}^2$, such that at each $\rho, \in \mathbb{R}, E > E_H$:

(1) There is exactly one point $Z_H(\rho,E) \in \mathscr{D}_H$ with the properties that for $\zeta = Z_H(\rho,E)$ the solution $x_i^H(t;z,E)$, $p_i^H(t;z,E)$ $(i = 1,2)$ exists for all $t \in \mathbb{R}$ and that $x_2^H(t;z,E)$ [and hence $p_2^H(t;z,E)$] is periodic in t with period $\tau_H((\rho,Z_H(\rho,E)),E)$.

(2) This periodic solution is not orbitally stable if $A_2 < 0$.

(3) Writing $Z_H(\rho,E) = (X_H(\rho,E),P_H(\rho,E))$, X_H, P_H are analytic functions of ρ, E at each such ρ, E and have the asymptotic behavior

$$X_H(\rho,E) = O(E^{-1}), \quad H = H_{\rm NR}, H_{\rm R}, \tag{2.8a}$$

$$P_H(\rho,E) = \begin{cases} (2E)^{-1/2}h(\rho) + O(E^{-1}), & H = H_{\rm NR}, \\ h(\rho) + O(E^{-1}), & H = H_{\rm R}, \end{cases} \tag{2.8b}$$

for $E \to \infty$, uniformly with respect to ρ on $\mathbb{R}$, where

$$h(\rho) = -\int_\rho^{\rho+1} u \frac{\partial V(u,0)}{\partial x_2} du. \tag{2.9}$$

Remark: Assertion (2) is intuitively obvious. By the Hartman–Grobner theorem,[11] in order to prove this statement it essentially suffices to establish the hyperbolicity of the appropriate fixed point of the maps $\mathscr{P}_{\rm NR}(\rho,\epsilon)$, $\mathscr{P}_{\rm R}(\rho,\epsilon)$ defined in Secs. III B and IV B, respectively (see also Sec. II B). On the other hand, the proof of the orbital stability property asserted in Theorem 2 for the case when $A_2 > 0$ is much more involved, since the stability of solutions of nonlinear Hamiltonian systems cannot in general be proved on the basis of linearized treatments.

We will denote by $\tilde{x}_i^H(t;\rho,E)$, $\tilde{p}_i^H(t;\rho,E)$ $(i = 1,2)$ the unique solution of (2.5) in assertion (1) of Theorem 1 satisfying (2.6) with $\zeta = Z_H(\rho,E)$.

Theorem 2: Let V satisfy conditions (I)–(III). Then there exists a constant $E'_H \geqslant E_H$ such that the solution $\tilde{x}_i^H(t;\rho,E)$, $\tilde{p}_i^H(t;\rho,E)$ $(i = 1,2)$ is orbitally stable for $\rho \in \mathbb{R}$, $E > E'_H$.

B. Strategy of proof

We will only explain it for the NR model, since the proofs for the R model are similar. All subsequent remarks in this subsection refer to the case $H = H_{\rm NR}$.

In this case, we denote $x_i^H(t;z,\epsilon)$, $p_i^H(t;z,\epsilon)$ by $x_i(t;z,\epsilon)$, $p_i(t;z,\epsilon)$, respectively, for $E > 0$, where ϵ, defined at each such E by $(2E)^{-1/2}$ for the NR model, is a more convenient variable than E for the present and later discussions. Introducing the new independent variable $u = x_1(t;z,\epsilon)$, one can perform an isoenergetic reduction[12] of system (2.5) for $H = H_{\rm NR}$ to a system of two first-order nonautonomous Hamiltonian differential equations satisfied by $X(u;z,\epsilon) = x_2(t;z,\epsilon)$, $P(u;z,\epsilon) = p_2(t;z,\epsilon)$. In this paper we proceed oppositely. We define $X(u;z,\epsilon)$, $P(u;z,\epsilon)$ as solutions of this nonautonomous system obeying appropriate initial conditions and express the solution $x_i(t;z,\epsilon)$, $p_i(t;z,\epsilon)$ $(i = 1,2)$ in terms of the first solution and the inverse mapping $u \mapsto t$, which exists for sufficiently small $\epsilon > 0$ and suitable u,z. In contrast to the latter solution of (2.5) for $H = H_{\rm NR}$, the functions X,P are analytic at $\epsilon = 0$ for appropriate u,z, and hence can be expanded in power series in ϵ having nonzero radii of convergence at these u,z. The coefficients in these power series can be evaluated systematically with relatively modest labor by using the integral equations for X,P which are equivalent to the corresponding differential equations plus the pertinent initial conditions.

Given that V satisfies (I) and is one-periodic in x_1, it is natural to define a canonical map[13] $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ which sends points $(\xi,\eta) \in \mathbb{R}^2$ to $(\xi' = X(\rho + 1,z,\epsilon), \eta' = P(\rho + 1;z,\epsilon))$. This map is well defined for (ξ,η) sufficiently close to the origin, provided that $\epsilon \geqslant 0$ is small enough.

Our strategy for proving Theorem 1 for $H = H_{\rm NR}$ in Sec. III C is very straightforward. An essential step in the proof is to show that under the hypotheses of the theorem $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ has exactly one fixed point $Z_1(\rho,\epsilon) = (X_1(\rho,\epsilon), P_1(\rho,\epsilon))$ in a certain ρ,ϵ-independent neighborhood of the origin for $\epsilon \downarrow 0$. In the notation of Theorem 1, $Z_H(\rho,E) = Z_1(\rho,\epsilon)$, $X_H(\rho,E) = X_1(\rho,\epsilon)$, $P_H(\rho,E) = P_1(\rho,\epsilon)$ for $E > 0$, $H = H_{\rm NR}$. The existence and uniqueness of $Z_1(\rho,\epsilon)$ follow by applying Theorem A.1 of Appendix A, while the desired analyticity properties of X_1, P_1, for small enough $\epsilon \geqslant 0$ follow by using Theorem A.2 of that appendix. From this and the existence and properties of the diffeomorphism $u \mapsto t$, we infer the existence of the unique solution $\tilde{x}_i^H(t;\rho,E)$, $\tilde{p}_i^H(t;\rho,E)$ $(i = 1,2)$ at the stated ρ, E values. The orbital instability of this solution for $A_2 < 0$ is proved by showing that $Z_1(\rho,\epsilon)$ is a hyperbolic fixed point at sufficiently small positive ϵ. Finally, Eqs. (2.8a) and (2.8b) are proved very easily for the present case $H = H_{\rm NR}$ by using, in particular, the above power-series expansion formulas for X and P.

To prove Theorem 2 for $H = H_{\rm NR}$, we reduce $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ to Birkhoff normal form. To accomplish this reduction, we first define a map $\mathscr{P}_{\rho,\epsilon}$ conjugate to $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ and which exists for $\rho \in \mathbb{R}$, $0 < \epsilon \leqslant \epsilon_3$, where ϵ_3 is a sufficiently small positive constant. Indeed, at each $\rho \in \mathbb{R}$, $0 < \epsilon \leqslant \epsilon_3$, this map has an elliptic fixed point at $\zeta = 0$ corresponding to the elliptic fixed point of $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ at $\zeta = Z_1(\rho,\epsilon)$, the Floquet multipliers of $\mathscr{P}_{\rho,\epsilon}$ at $\zeta = 0$ are nonreal, of the form $\lambda(\rho,\epsilon)$, $\overline{\lambda(\rho,\epsilon)}$, unimodular, and satisfy certain nonresonance conditions, and this map has certain analyticity properties. These facts entail that at the latter ρ,ϵ values $\mathscr{P}_{\rho,\epsilon}$, and hence $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ is conjugate to a map $(z,\bar{z}) \to (z^*,\bar{z}^*)$ in Birkhoff normal form:

$$z^* = \lambda(\rho,\epsilon)z[1 + \mu(\rho,\epsilon)|z|^2] + O(|z|^4), \quad z \to 0, \tag{2.10}$$

where z is a complex variable. Moreover, the first twist coefficient $\mu(\rho,\epsilon)$ is nonzero for $\rho \in \mathbb{R}$, $0 < \epsilon \leqslant \epsilon_3$.

These properties of $\mathscr{P}_{\rho,\epsilon}$ were derived by estimating sufficiently accurately in the limit $\epsilon \downarrow 0$ the derivatives of ξ',η' with respect to ξ,η of orders $1 \leqslant n \leqslant 3$ at $\zeta = Z_1(\rho,\epsilon)$. These estimates were readily obtained by differentiating the pertinent coefficients (obtained by the above systematic procedure) of the power series in ϵ of X,P and by using a crude estimate for $Z_1(\rho,\epsilon)$ in the latter limit.

The stated attributes of the map (2.10) and analyticity and domain properties thereof allow us to apply the version of the twist theorem in Appendix B to it. By this application and simple additional arguments, we infer that the orbital stability property asserted by Theorem 2 holds under the stated conditions for the case $H = H_{\rm NR}$.

III. PROOF OF THEOREMS 1 AND 2 FOR THE NR MODEL

It is mathematically natural and convenient to replace E in this paper by a suitable parameter ϵ, which in this section is defined by

$$\epsilon = (2E)^{-1/2} \tag{3.1}$$

for $E > 0$. In discussing the analyticity of certain functions, we will allow ϵ to vanish.

This section consists of Secs. III A–III C. In Sec. III A, we collect certain existence, uniqueness, and analyticity results and state certain integral equations that play a central role in Secs. III and IV. In Sec. III B, we define more precisely the canonical map $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ introduced in Sec. II B and we derive an asymptotic formula for the image $\mathscr{P}_{\rm NR}(\rho,\epsilon) \times (\xi,\eta)$ of points (ξ,η) near the origin for $\epsilon \downarrow 0$. The latter formula is essential in the proof of Eqs. (2.8) for the case $H = H_{\rm NR}$ in Sec. III C, where we prove Theorem 1 for this case.

A. Auxiliary considerations

Our only assumptions on $V(x_1,x_2)$ in the present subsection is that it has property (I) and is one-periodic in x_1 in the strip (2.1).

As in Sec. II B we denote the solution $x_i^H(t;z,E)$, $p_i^H(t;z,E)$ $(i = 1,2)$ defined in Sec. II A by $x_i(t;z,\epsilon)$, $p_i(t;z,\epsilon)$ in the case $H = H_{\rm NR}$ for $E > 0$.

We will not work directly with this solution, but rather with the functions

$$X(u;z,\epsilon) = x_2(t;z,\epsilon)\,, \tag{3.2a}$$

$$P(u;z,\epsilon) = p_2(t;z,\epsilon)\,, \tag{3.2b}$$

where u is a new independent variable given by

$$u = x_1(t;z,\epsilon)\,. \tag{3.2c}$$

By (2.6) and (3.2),

$$X(\rho;z,\epsilon) = \xi\,, \tag{3.3a}$$

$$P(\rho;z,\epsilon) = \eta\,. \tag{3.3b}$$

By the definition of the solution $x_i(t;z,\epsilon)$, $p_i(t;z,\epsilon)$ $(i = 1,2)$ [and particularly the condition $p_1(t;z,\epsilon) > 0$] together with (3.2), if follows that at those $z\in\mathbb{R}^3$, $\epsilon > 0$ with $(\rho,\xi)\in\mathscr{S}$ [see (2.1)] at which this solution exists, $X(u;z,\epsilon)$, $P(u;z,\epsilon)$ satisfy the isoenergetically reduced system of Hamiltonian differential equations corresponding to (2.5) for $H = H_{\rm NR}$:

$$\frac{dX}{du} = \epsilon\,\frac{P}{\sigma(u,Z,\epsilon)}\,, \tag{3.4a}$$

$$\frac{dP}{du} = \epsilon\,\frac{f(u,X)}{\sigma(u,Z,\epsilon)}\,, \tag{3.4b}$$

where $Z = (X,P)$ and

$$\sigma(u,w,\epsilon) = \{1 - \epsilon^2[2V(u,w_1) + w_2^2]\}^{1/2}\,, \tag{3.5a}$$

$$f(x_1,x_2) = -\frac{\partial V(x_1,x_2)}{\partial x_2}\,, \tag{3.5b}$$

with $w = (w_1,w_2)$. Notice that the denominators in (3.4) have the property that $\sigma(u,Z(u;z,\epsilon),\epsilon) = \epsilon p_1(t;z,\epsilon) > 0$ when $x_i(t;z,\epsilon), p_i(t;z,\epsilon)$ $(i = 1,2)$ exist at given $t,z,\epsilon > 0$ values. Here $Z(u;z,\epsilon) = (X(u;z,\epsilon), Y(u;z,\epsilon))$.

The integral equations

$$X(u;z,\epsilon) = \xi + \epsilon \int_\rho^u \frac{P(u';z,\epsilon)}{\sigma(u',Z(u';z,\epsilon),\epsilon)}\,du'\,, \tag{3.6a}$$

$$P(u;z,\epsilon) = \eta + \epsilon \int_\rho^u \frac{f(u',X(u';z,\epsilon))}{\sigma(u',Z(u';z,\epsilon),\epsilon)}\,du'\,, \tag{3.6b}$$

which incorporate conditions (3.3), will be very useful in this section.

Instead of deriving the relevant properties of $X(u;z,\epsilon)$, $P(u;z,\epsilon)$ from those of $x_i(t;z,\epsilon), p_i(t;z,\epsilon)$ $(i = 1,2)$ as in Ref. 1, it is more convenient here to proceed in the reverse order. Let V possess property (I). Then at each $z\in\mathbb{R}^3$, $\epsilon \geqslant 0$ with $(\rho,\xi)\in\mathscr{S}$ at which it exists, we henceforth *define* $X(u;z,\epsilon)$, $P(u;z,\epsilon)$ as a solution of (3.4) at the points u in a maximum interval in $\mathbb{R}$ containing $u = \rho$, which satisfies the conditions (3.3) and is such that $(u,X(u;z,\epsilon))\in\mathscr{S}$ and $\sigma(u;Z(u;z,\epsilon),\epsilon) > 0$ at each u in this interval. By this definition and the smoothness properties of the right sides of Eqs. (3.4) under the present hypothesis that property (I) of V obtains, it follows that this solution is unique.

We will need the following version of the usual theorems on existence and analyticity of solution of ordinary differential equations.

Lemma 3.1: Let V obey (I) and be one-periodic in the strip (2.1). Choose positive constants a, b with $a < a_0$, $b < b_0$, a_0 being as in (2.1) and b_0 arbitrary. There exists a positive constant ϵ_0 such that $X(u;z,\epsilon)$, $P(u;z,\epsilon)$ exist for $u\in J(\rho,\epsilon)$, $z = (\rho,\zeta)\in\mathscr{V}$, $0 \leqslant \epsilon \leqslant \epsilon_0$. In particular, ϵ_0 is small enough so that $\sigma(u,w,\epsilon) > c > 0$, where c is a constant; $J(\rho,\epsilon) = [\rho - C/\epsilon, \rho + C/\epsilon] \supset [\rho,\rho + 1]$ for $\rho\in\mathbb{R}$, $0 < \epsilon \leqslant \epsilon_0$, and $J(\rho,0) = \mathbb{R}$ for $\rho\in\mathbb{R}$, C being a constant; and

$$\mathscr{V} = \mathbb{R}\times\mathscr{U}\,, \tag{3.7a}$$

$$\mathscr{U} = \{(u_1,u_2)\in\mathbb{R}^2\colon\ |u_1| \leqslant a,\ |u_2| \leqslant b\}\,. \tag{3.7b}$$

Moreover, at each such u,z,ϵ the functions $X(u;z,\epsilon)$ $P(u;z,\epsilon)$ are real analytic in these variables and such that

$$(X(u;z,\epsilon),P(u;z,\epsilon))\in\mathscr{U}_0\,,$$

where $\mathscr{U}_0$ is $\mathscr{U}$ with a,b replaced by a_0, b_0, respectively.

Remark: The analyticity of $X(u;z,\epsilon)$, $P(u;z,\epsilon)$ at $\epsilon = 0$ at the relevant u,z, values will be basic in what follows. The fact that these two functions can be continued analytically to suitable negative ϵ values is of no interest here.

Proof of Lemma 3.1: It is based on the fact, implied by (3.5) and our hypothesis on V, that the right sides of (3.4) are real analytic in their arguments for $(u,Z,\epsilon)\in\mathbb{R}\times\mathscr{U}\times[0,\epsilon_0]$ and one-periodic there, provided that ϵ_0 is small enough, and on applying the usual successive-approximation method to (3.6) in the complex domain. ■

We now define a mapping which is the inverse of the formally defined mapping (3.2c). At each $z\in\mathscr{V}$, $0 < \epsilon \leqslant \epsilon_0$, let

$$u \mapsto t \tag{3.8a}$$

be the diffeomorphism from $J(\rho,\epsilon)$ onto a closed interval in $\mathbb{R}$, with

$$t = \epsilon \int_\rho^u \frac{1}{\sigma(u',Z(u';z,\epsilon),\epsilon)} du' = g(u;z,\epsilon) . \tag{3.8b}$$

That this is a diffeomorphism follows since $g(u;z,\epsilon)$ is strictly increasing in u on $J(\rho,\epsilon)$ at each such z,ϵ and is real analytic in its arguments at these u,z,ϵ, and by invoking the implicit function theorem. These properties of g follow by (3.5a), (3.8b), and Lemma 3.1. Hence $g^{-1}(t;z,\epsilon)$ (g^{-1} = inverse of g) is analytic in its arguments for $(t,z,\epsilon) \in g(J(\rho,\epsilon);z,\epsilon)\times \mathscr{V} \times (0,\epsilon_0]$ ($g(J(\rho,\epsilon);z,\epsilon)$ = image of $J(\rho,\epsilon)$ under $u \mapsto t$ at given z,ϵ).

At the latter t,z,ϵ, define $x_1(t;z,\epsilon) = g^{-1}(t;z,\epsilon)$, $p_1(t;z,\epsilon) = \dot{x}_1(t;z,\epsilon)$, and define $x_2(t;z,\epsilon)$, $p_2(t;z,\epsilon)$ there by (3.2a) and (3.2b). Then $x_i(t;z,\epsilon)$, $p_i(t;z,\epsilon)$ $(i = 1,2)$ exist and constitute the unique solution of (2.5) satisfying (2.6) at each such t,z,ϵ. Moreover, these four functions are analytic in their arguments at these t,z,ϵ values and agree there with the respective functions defined earlier in this subsection. These results follow easily, in particular by Lemma 3.1 and the stated properties of the diffeomorphism (3.8a).

B. Definition of the map $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ and formulas for this map for $\epsilon \downarrow 0$

In this subsection, we assume that V has properties (I) and (II).

Definition: For each $\rho \in \mathbb{R}$, $0 \leqslant \epsilon \leqslant \epsilon_0$, $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ is a mapping with domain $\mathscr{U}$ [see (3.7b)] which sends each $(\xi,\eta) \in \mathscr{U}$ into

$$(\xi',\eta') = (X(\rho + 1,z,\epsilon), P(\rho + 1,z,\epsilon)) \in \mathbb{R}^2 .$$

Remark: $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ is well defined and is a (local) canonical map.[13] Indeed, ξ', η' are analytic in z,ϵ for $(z,\epsilon) \in \mathscr{V} \times [0,\epsilon_0]$ and one-periodic in ρ there. These analyticity and periodicity properties follow by the analyticity attributes of $X(u;z,\epsilon)$, $P(u;z,\epsilon)$ stated in Lemma 3.1 and the fact that these two functions are invariant under the substitutions $u \to u + 1$, $\rho \to \rho + 1$ for $u \in [\rho,\rho + 1]$, $(z,\epsilon) \in \mathscr{V} \times [0,\epsilon_0]$ [see the sentence of Appendix A containing (A12)]; and the canonical nature of $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ by, in particular, the Hamiltonian character of Eqs. (3.4) and a well-known theorem.[14]

The next lemma will be of central importance in this section.

Lemma 3.2: For $z \in \mathscr{V}$,

$$X(\rho + 1;z,\epsilon) = \xi + \epsilon\eta + \epsilon^2 \int_\rho^{\rho+1} (\rho + 1 - u) f(u,\xi) du + \epsilon^3\left[\eta g(\rho,\xi) + \frac{\eta^3}{4}\right] + O(\epsilon^4) , \tag{3.9a}$$

$$P(\rho + 1;z,\epsilon) = \eta + \epsilon \int_\rho^{\rho+1} f(u,\xi) du + \epsilon^2 \eta \int_\rho^{\rho+1} (u - \rho) f'(u,\xi) du + \epsilon^3 [h(\rho,\xi) + \eta^2 j(\rho,\xi)] + O(\epsilon^4) , \tag{3.9b}$$

as $\epsilon \downarrow 0$, uniformly in z, where $f'(u,\xi) = \partial f(u,\xi)/\partial\xi$, and where g,h,j are bounded analytic functions of ρ,ξ for $\rho \in \mathbb{R}$, $|\xi| \leqslant a$ (a is as Lemma 3.1) which are η,ϵ-independent.

Remarks: (1) In what follows, equations in which the symbols ρ and/or z appear should be regarded as holding for $\rho \in \mathbb{R}$ and/or $z \in \mathscr{V}$, respectively, and estimates involving $O(\epsilon^n)$ symbols should be understood to hold for $\epsilon \downarrow 0$, uniformly with respect to the remaining variables, even if this is not mentioned explicitly.

(2) The obvious reason why the limit $\epsilon \downarrow 0$, rather than the two-sided limit $\epsilon \to 0$ occurs in this paper is that we did not choose to extend the relevant functions to negative values of ϵ. We stress that this is a question of personal taste which is of no mathematical importance here.

Proof of Lemma 3.2: Since $X(u;z,\epsilon)$, $P(u;z,\epsilon)$ are analytic in ϵ at $\epsilon = 0$ for $u \in [\rho,\rho + 1]$, $z \in \mathscr{V}$, we can expand these functions in convergent power series in ϵ at such u,z. Inserting these series in (3.6a) and (3.6b) and equating coefficients of equal powers of ϵ, one sees that the coefficients can be calculated inductively. The coefficients of the powers ϵ^r $(r \leqslant 3)$ in (3.9a) and (3.9b) were calculated in this way with relatively little labor. The stated properties of g,h,j follow from the explicit forms of these coefficients (which we omit) and analogous properties of V.

Consider the Taylor series in ϵ of $X(\rho + 1;z,\epsilon)$, $P(\rho + 1;z,\epsilon)$ with $O(\epsilon^n)$ remainders. Since $X(\rho + 1;z,\epsilon)$, $P(\rho + 1;z,\epsilon)$ are analytic in z,ϵ and one-periodic in ρ if $(z,\epsilon) \in \mathscr{V} \times [0,\epsilon_0]$, and since the periodicity region $[0,1] \times \mathscr{U} \times [0,\epsilon_0]$ is compact, they are also bounded over $\mathscr{V} \times [0,\epsilon_0]$. Therefore each such remainder is the product of ϵ^n by a function of z,ϵ analytic in these variables over $\mathscr{V} \times [0,\epsilon_0]$ and bounded there, whence the uniformity of the estimates (3.9) in z follows. ∎

For later use, we record the following equations, entailed by (3.9) in view of the stated properties of g,h,j, the analyticity and boundedness of $f(u,\xi)$ for $u \in \mathbb{R}$, $|\xi| \leqslant a$, the boundedness of $\mathscr{U}$ [see (3.7b)], and the structure of the terms $O(\epsilon^4)$ in (3.9) according to the last paragraph:

$$\frac{\partial X(\rho + 1;z,\epsilon)}{\partial\xi} = 1 + \epsilon^2 \int_\rho^{\rho+1} (\rho + 1 - u) f'(u,\xi) du + \eta O(\epsilon^3) + O(\epsilon^4) , \tag{3.10a}$$

$$\frac{\partial X(\rho + 1;z,\epsilon)}{\partial\eta} = \epsilon + O(\epsilon^3) , \tag{3.10b}$$

$$\frac{\partial P(\rho + 1;z,\epsilon)}{\partial\xi} = \epsilon \int_\rho^{\rho+1} f'(u,\xi) du + \eta O(\epsilon) + O(\epsilon^3) , \tag{3.10c}$$

$$\frac{\partial P(\rho + 1;z,\epsilon)}{\partial\eta} = 1 + \epsilon^2 \int_\rho^{\rho+1} (u - \rho) f'(u,\xi) du + \eta O(\epsilon^3) + O(\epsilon^4) . \tag{3.10d}$$

C. Proof of Theorem 1 for $H = H_{\rm NR}$

In this subsection, we assume that V satisfies conditions (I), (II), and that

$$A_2 \neq 0 . \tag{3.11}$$

We begin by stating and proving Theorem 3.1. As explained in the next Remark, this theorem and simple additional arguments entail the assertions of Theorem 1 for the case $H=H_{\rm NR}$, except that it furnishes a rougher estimate of $Z_H(\rho,E)$ in the limit $\epsilon\downarrow 0$ for this case than do Eqs. (2.8). For $H=H_{\rm NR}, E>0$, we write $Z_1(\rho,\epsilon)=Z_H(\rho,\epsilon)$. The derivation of the results (2.8) pertaining to the NR model will be given in this section after the proof of the next theorem.

Theorem 3.1: Let V have properties (I), (II), and (3.11). Then:

(1) There exist positive constants $\alpha\leqslant\min\{a,b\}$, $\epsilon_1\leqslant\epsilon_0$ with a, b as in Lemma 3.1, such that for $0<\epsilon\leqslant\epsilon_1$ there is exactly one solution $X_1(u,\epsilon)$, $P_1(u,\epsilon)$ of Eqs. (3.5) which exists for all $u\in\mathbb{R}$, is one-periodic in u, and has the property that $Z_1(u,\epsilon)=(X_1(u,\epsilon), P_1(u,\epsilon))$ has norm $\|Z_1(u,\epsilon)\|\leqslant\alpha$ at all such u. [Note that $Z_1(\rho,\epsilon)\in\mathscr{U}$ at these ρ, ϵ.]

(2) The estimate

$$\|Z_1(u,\epsilon)\|=O(\epsilon) \tag{3.12}$$

holds uniformly with respect to u on $\mathbb{R}$ in the limit $\epsilon\downarrow 0$.

(3) At each $0<\epsilon\leqslant\epsilon_1$, the above periodic solution is elliptic (resp. ordinary hyperbolic) if $A_2>0$ (resp. $A_2<0$).

(4) $X_1(u,\epsilon)$, $P_1(u,\epsilon)$ are real analytic functions of their arguments for $u\in\mathbb{R}$, $0<\epsilon\leqslant\epsilon_1$ and have real analytic extensions to $u\in\mathbb{R}$, $0\leqslant\epsilon\leqslant\epsilon_1$, which will be denoted by the same respective symbols.

Remark: Theorem 1 for $H=H_{\rm NR}$, except for the indicated estimate, follows easily from Theorem 3.1, supplemented in particular by Lemma 3.1 and the fact that (3.8b), with $Z(u';z,\epsilon)$ replaced by $Z_1(u',\epsilon)$, defines a global diffeomorphism $u\mapsto t$ from $\mathbb{R}$ onto $\mathbb{R}$ at each $(\rho,\epsilon)\in\mathbb{R}\times(0,\epsilon_1]$. Note especially that $\tilde{x}_2^H(t;\rho,E)=X_1(u,\epsilon)$, $\tilde{p}_2^H(t;\rho,E)=P_1(u,\epsilon)$ for the NR model if $(\rho,u,\epsilon)\in\mathbb{R}\times\mathbb{R}\times(0,\epsilon_1]$, where t is given by this global diffeomorphism.

Proof of Theorem 3.1: It will be effected in three major steps: (a) proof of (1) and (2); (b) proof of (3); (c) proof of (4).

(a) To prove existence and uniqueness we write Eqs. (3.4) in the form

$$\frac{dZ}{du}=\epsilon\phi(u,Z,\epsilon)\,, \tag{3.13a}$$

where $Z=(X,P)$ as before and

$$\phi(u,w,\epsilon)=(w_2/\sigma(u,w,\epsilon),f(u,w_1)/\sigma(u,w,\epsilon))\,, \tag{3.13b}$$

where $w=(w_1,w_2)$ as previously. Hence

$$\int_0^1\phi(u,0,0)du=0\,, \tag{3.14a}$$

$$\int_0^1\frac{\partial\phi(u,0,0)}{\partial w}du=\begin{bmatrix}0 & 1\\ -A_2 & 0\end{bmatrix}\equiv N\,, \tag{3.14b}$$

by (2.2), (2.3), (3.5), and (3.13b). In view of (3.13a), (3.14a), and the facts that N in (3.14b) in nonsingular and that $\phi(u,w,\epsilon)$ is one-periodic in u and real analytic in u,w,ϵ in a neighborhood of the subset $\{(u,w=0,\epsilon=0),\ u\in\mathbb{R}\}\subset\mathbb{R}^4$, we can apply Theorem A.1 of Appendix A and the remark thereto to prove assertions (1) and (2).

(b) Let $M_1(\rho,\epsilon)$ denote the Jacobian matrix $\partial\mathscr{P}_{\rm NR}(\rho,\epsilon)(\zeta)/\partial\zeta$ of the map $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ at its fixed point $\zeta=Z_1(\rho,\epsilon)$. By, in particular, the definition of this map together with (3.10a), (3.10d), (3.12), the fact that

$$f'(u,\xi)=f'(u,0)+\xi f''(u,\theta\xi) \tag{3.15}$$

at the u,ξ of interest, where $\theta=\theta(u,\xi)\in(0,1)$, the boundedness of $f''(u,\xi)=\partial^2f(u,\xi)/\partial\xi^2$ for $u\in\mathbb{R}$, $|\xi|\leqslant a$, and (2.3) and (3.5b), we obtain

$$\operatorname{Tr}M_1(\rho,\epsilon)=2-A_2\epsilon^2+O(\epsilon^3) \tag{3.16}$$

for $\rho\in\mathbb{R}$. Whence we can assume without loss of generality that the constant $\epsilon_1>0$ in assertion (1) is so small that for $(\rho,\epsilon)\in\mathbb{R}\times(0,\epsilon_1]$ we have $0<\operatorname{Tr}M_1(\rho,\epsilon)<2$ if $A_2<0$ and $\operatorname{Tr}M_1(\rho,\epsilon)>2$ if $A_2>0$. Since $\det M_1(\rho,\epsilon)=1$ at each such ρ,ϵ, $Z_1(\rho,\epsilon)$ is an elliptic (resp. ordinary hyperbolic) fixed point of $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ at these ρ,ϵ values when $A_2>0$ (resp. $A_2<0$). Hence (3) has been proved.

(c) Specializing Theorem A.2 of Appendix A to the present situation, (4) follows. ■

By the definition of $\mathscr{P}_{\rm NR}(\rho,\epsilon)$, $\zeta\in\mathscr{U}$ is a fixed point of $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ at given $\rho\in\mathbb{R}$, $0\leqslant\epsilon\leqslant\epsilon_1$ iff

$$X(\rho+1;z,\epsilon)=\xi\,, \tag{3.17a}$$

$$P(\rho+1;z,\epsilon)=\eta\,. \tag{3.17b}$$

These equations yield each coefficient in the power series expansion in ϵ of $Z_1(\rho,\epsilon)$ in terms of coefficients of the relevant series for X,P. Here our goal is the more modest one of showing that $X_1(\rho,\epsilon)$, $P_1(\rho,\epsilon)$ are given by the right sides of (2.8a), (2.8b), respectively, appropriate to the case $H=H_{\rm NR}$. This will be done by specializing (3.9a) and (3.17a) to $\zeta=Z_1(\rho,\epsilon)$, dividing by ϵ, applying the mean-value theorem analogously to how it was applied in (3.15), and using the pertinent boundedness properties of f' and g as well as (II), (2.9), (3.5b), and (3.12), thus obtaining

$$\begin{aligned}P_1(\rho,\epsilon)&=-\epsilon\int_\rho^{\rho+1}(\rho+1-u)f(u,X_1(\rho,\epsilon))du+O(\epsilon^3)\\&=\epsilon h(\rho)+O(\epsilon^2)\,.\end{aligned} \tag{3.18a}$$

Similar arguments, making use of (II), (2.4a), (3.5a), (3.12), and (3.17b) in particular, yield

$$X_1(\rho,\epsilon)=O(\epsilon^2)\,, \tag{3.18b}$$

and hence we have proved the desired asymptotic result for $Z_1(\rho,\epsilon)$.

D. Proof of Theorem 2 for $H=H_{\rm HR}$

In this subsection, V is assumed to possess properties (I)–(III).

Let $\rho\in\mathbb{R}$, $0\leqslant\epsilon\leqslant\epsilon_1$. We define $\mathscr{P}_{\rho,\epsilon}$ as a map conjugate to $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ by a translation $\mathscr{J}_{\rho,\epsilon}:\zeta\mapsto\zeta-Z_1(\rho,\epsilon)$:

$$\mathscr{P}_{\rho,\epsilon}(\zeta)=\mathscr{J}_{\rho,\epsilon}^{-1}\circ\mathscr{P}_{\rm NR}(\rho,\epsilon)\circ\mathscr{J}_{\rho,\epsilon}(\zeta)\,, \tag{3.19}$$

for $\zeta\in\mathscr{U}-Z_1(\rho,\epsilon)$. Since $\zeta=Z_1(\rho,\epsilon)$ is a fixed point of $\mathscr{P}_{\rm NR}(\rho,\epsilon)$, $\zeta=0$ is a fixed point of $\mathscr{P}_{\rho,\epsilon}$. Moreover, writing $(\xi^*,\eta^*)=\mathscr{P}_{\rho,\epsilon}(\xi,\eta)$, we have

$$\xi^*=\sum_{n=0}^{\infty}\sum_{0\leqslant s\leqslant n}[(n-s)!s!]^{-1}\left(\frac{\partial^n\xi'}{\partial\xi^{n-s}\partial\eta^s}\right)_1\xi^{n-s}\eta^s\,, \tag{3.20a}$$

$$\eta^*=\sum_{n=0}^{\infty}\sum_{0\leqslant s\leqslant n}[(n-s)!s!]^{-1}\left(\frac{\partial^n\eta'}{\partial\xi^{n-s}\partial\eta^s}\right)_1\xi^{n-s}\eta^s\,, \tag{3.20b}$$

where at the latter ρ,ϵ the series converge for (ξ,η) in a ρ,ϵ-independent rectangle centered at $\zeta=0$, where

$\xi' = X(\rho+1;z,\epsilon)$, $\eta' = P(\rho+1;z,\epsilon)$, as before, and where the subscript 1 means that the derivatives are evaluated at $\zeta = Z_1(\rho,\epsilon)$. The existence of such a rectangle follows from the facts that ξ^*,η^* are analytic functions of ρ,z,ϵ on $\mathbb{R}\times\mathscr{U}\times[0,\epsilon_1]$ and are one-periodic in ρ there, so that the periodicity region $[0,1]\times\mathscr{U}\times[0,\epsilon_1]$ is compact.

An essential step in the proof of Theorem 2 for $H = H_{\rm NR}$ is the reduction of $\mathscr{P}_{\rho,\epsilon}$ to Birkhoff normal form for small enough $\epsilon > 0$. In order to perform this reduction, one needs to know sufficiently accurately the behavior of the derivatives

$$\left(\frac{\partial^n \xi'}{\partial \xi^{n-s}\partial\eta^s}\right)_1, \quad \left(\frac{\partial^n \eta'}{\partial \xi^{n-s}\partial\eta^s}\right)_1 \quad (0\leqslant s\leqslant n),$$

for $1\leqslant n\leqslant 3$ in the limit $\epsilon\downarrow 0$. This behavior is dealt with in the next lemma.

Lemma 3.3: The following equations hold uniformly in ρ on $\mathbb{R}$:

$$\left(\frac{\partial\xi'}{\partial\xi}\right)_1 = 1 + \epsilon^2\int_\rho^{\rho+1}(1+\rho-u)f'(u,0)du + O(\epsilon^4), \tag{3.21a}$$

$$\left(\frac{\partial\xi'}{\partial\eta}\right)_1 = \epsilon + O(\epsilon^3), \tag{3.21b}$$

$$\left(\frac{\partial\eta'}{\partial\xi}\right)_1 = -\epsilon A_2 + O(\epsilon^3), \tag{3.21c}$$

$$\left(\frac{\partial\eta'}{\partial\eta}\right)_1 = 1 + \epsilon^2\int_\rho^{\rho+1}(u-p)f'(u,0)du + O(\epsilon^4), \tag{3.21d}$$

and for $n = 2,3$,

$$\left(\frac{\partial^n\xi'}{\partial\xi^{n-s}\partial\eta^s}\right)_1 = O(\epsilon^{2+s}), \quad 0\leqslant s\leqslant 2, \tag{3.22a}$$

$$\left(\frac{\partial^3\xi'}{\partial\eta^3}\right)_1 = O(\epsilon^3), \tag{3.22b}$$

$$\left(\frac{\partial^n\eta'}{\partial\xi^n}\right)_1 = -\epsilon A_n + O(\epsilon^3), \tag{3.22c}$$

$$\left(\frac{\partial^n\eta'}{\partial\xi^{n-s}\partial\eta^s}\right)_1 = O(\epsilon^{1+s}), \quad 1\leqslant s\leqslant n. \tag{3.22d}$$

Proof of Lemma 3.3: Equations (3.21) follow immediately by (3.10), (3.15), (3.18), and the one-periodicity of $V(x_1,x_2)$ in x_1. Similar considerations, using Lemma 3.2 and (3.18) in particular, yield (3.22).[15] ■

There exists a sufficiently small positive constant $\epsilon_2\leqslant\epsilon_1$ such that the eigenvalues (properly labeled) of the Jacobian $M_1(\rho,\epsilon)$ of the map $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ at $Z_1(\rho,\epsilon)$, and therefore of the map $\mathscr{P}_{\rho,\epsilon}$ at $\zeta = 0$ have the following properties for $\rho\in\mathbb{R}$, $0<\epsilon\leqslant\epsilon_2$: (i) they are of the form $\lambda(\rho,\epsilon)$, $\overline{\lambda(\rho,\epsilon)}$ with $\operatorname{Im}\lambda(\rho,\epsilon) > 0$, $|\lambda(\rho,\epsilon)| = 1$, and λ is an analytic function of ρ,ϵ at each such ρ,ϵ which is one-periodic in ρ; (ii) $\lambda^3(\rho,\epsilon), \lambda^4(\rho,\epsilon)\neq 1$.

In view of the fact that $Z_1(\rho,\epsilon)$ is an elliptic fixed point of $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ at the latter ρ,ϵ values [see part (b), proof of Theorem 3.1], we only have to prove the stated analyticity property of $\lambda(\rho,\epsilon)$ and property (ii). This can be done by arguments of the same type as those used in a similar connection in Ref. 1. Notice that

$$\lambda(\rho,\epsilon) = 1 + iA_2\epsilon + O(\epsilon^3). \tag{3.23}$$

Let $\rho\in\mathbb{R}$, $0<\epsilon\leqslant\epsilon_2$. By the theory of reduction to Birkhoff normal form,[16,17] the facts that at such ρ,ϵ the map $\mathscr{P}_{\rho,\epsilon}$ is canonical and that $\mathscr{P}_{\rho,\epsilon}(\zeta)$ is analytic in ρ,ϵ,ζ and one-periodic in ρ for ζ in a ρ,ϵ-independent neighborhood of the origin, properties (i) and (ii) of $\lambda(\rho,\epsilon)$, (3.21) and (3.22), and additional considerations of a straightforward type, $\mathscr{P}_{\rho,\epsilon}$ is conjugate to a canonical mapping $\mathscr{Q}_{\rho,\epsilon}$ of the latter form at every such ρ,ϵ:

$$\mathscr{Q}_{\rho,\epsilon} = \phi_{\rho,\epsilon}^{-1}\circ\mathscr{P}_{\rho,\epsilon}\circ\phi_{\rho,\epsilon}|B_{\rho,\epsilon}. \tag{3.24}$$

At the latter ρ,ϵ, $\phi_{\rho,\epsilon}$ is a diffeomorphism from an open ball $B_{\rho,\epsilon} = \{(x,y)\in\mathbb{R}^2\colon \|(x,y)\| < r_{\rho,\epsilon}\}$, $r_{\rho,\epsilon} > 0$, onto an open neighborhood of $0\in\mathbb{R}^2$ and is one-periodic in ρ there. In more detail, $\phi_{\rho,\epsilon}(x,y) = (\phi_1(x,y;\rho,\epsilon), \phi_2(x,y;\rho,\epsilon))$, $\phi_{\rho,\epsilon}^{-1}(x',y') = (\psi_1(x',y';\rho,\epsilon), \psi_2(x',y';\rho,\epsilon))$, where ϕ_i,ψ_i $(i = 1,2)$ are analytic in their arguments for $\rho\in\mathbb{R}$, $0<\epsilon\leqslant\epsilon_2$, $(x,y)\in B_{\rho,\epsilon}$, $(x',y')\in\phi_{\rho,\epsilon}(B_{\rho,\epsilon})$. By these analyticity and periodicity properties of $\phi_{\rho,\epsilon}$, it follows that for each $\rho\in\mathbb{R}$ and ϵ in a compact subset of $(0,\epsilon_2]$ this function is defined on a ρ,ϵ-independent neighborhood for the origin. So is $\mathscr{Q}_{\rho,\epsilon}$ at each such ρ,ϵ because of this domain property of $\phi_{\rho,\epsilon}$, the fact that $\mathscr{P}_{\rho,\epsilon}$ has the same property, and (3.24). Write

$$(x^*,y^*) = \mathscr{Q}_{\rho,\epsilon}(x,y), \quad z = x+iy, \quad z^* = x^*+iy^*. \tag{3.25a}$$

Then

$$z^* = \lambda(\rho,\epsilon)z[1+\mu(\rho,\epsilon)|z|^2] + O(|z|^4), \quad z\to 0, \tag{3.25b}$$

at each such ρ,ϵ. Here $\mu(\rho,\epsilon)$, the first twist coefficient of $\mathscr{Q}_{\rho,\epsilon}$, is real-valued at the latter ρ,ϵ.

This twist coefficient depends on how the matrix $C(\rho,\epsilon)$ diagonalizing $M_1(\rho,\epsilon) = \partial P_{\rm NR}(\rho,\epsilon)(Z_1(\rho,\epsilon))/\partial\zeta$ for $\rho\in\mathbb{R}$, $0<\epsilon\leqslant\epsilon_2$, i.e.,

$$C(\rho,\epsilon)^{-1}M_1(\rho,\epsilon)C(\rho,\epsilon) = \begin{pmatrix}\lambda(\rho,\epsilon) & 0\\ 0 & \overline{\lambda(\rho,\epsilon)}\end{pmatrix}, \tag{3.26}$$

is normalized. We can (and will) select $C(\rho,\epsilon)$ so that at each ρ,ϵ its entries are analytic in these variables, one-periodic in ρ, and normalized as was the corresponding matrix $C(E)$ in Ref. 1 [see Eqs. (3.33) and (3.34) therein]. Then $\mu(\rho,\epsilon)$ has these same analyticity and periodicity properties at the latter ρ,ϵ and

$$\mu(\rho,\epsilon) = \tfrac{1}{8}A_2^{-2}(A_2A_4 - \tfrac{5}{3}A_3^2) + O(\epsilon^2) \tag{3.27}$$

for $\epsilon\downarrow 0$, uniformly in ρ on $\mathbb{R}$. Except for this uniformity, (3.27) is the same result obtained in Ref. 1 [see (3.31) therein] for the first twist coefficient pertaining to a map [called $\mathscr{P}_{\rm NR}(E)$ there] which is $\mathscr{P}_{\rho,\epsilon}$ specialized to the case $Z_1(\rho,\epsilon)\equiv 0$. This should be clear from the following facts: (1) the derivatives in Eqs. (3.21) and (3.22), when evaluated at $\zeta = 0$ rather than at $\zeta = Z_1(\rho,\epsilon)$, are also given by the right sides of these equations, which leads to results for $\lambda(\rho,\epsilon)$ [see (3.23)] and $C(\rho,\epsilon)$ which are the same, except for uniformity, as (3.23) and (3.35) of Ref. 1, respectively, to within the desired accuracy; (2) Eq. (3.30) of Ref. 1 understood in the present context. By this version of the latter equation and the facts that $\lambda(\rho,\epsilon)$ and $C(\rho,\epsilon)$, as chosen here, have extensions that are real analytic in ρ,ϵ and one-periodic in ρ for $\rho\in\mathbb{R}$, $0\leqslant\epsilon\leqslant\epsilon_2$, $\mu(\rho,\epsilon)$ has the same properties, whence the stated unformity of (3.27) follows by an argument of the same type as one used to prove that of the estimates (3.9).

Proof of Theorem 2 for $H = H_{\rm NR}$: By (2.4) and (3.27), there exists a positive constant $\epsilon_3\leqslant\epsilon_2$ such that $\mu(\rho,\epsilon)\neq 0$ for $\rho\in\mathbb{R}$, $0<\epsilon\leqslant\epsilon_3$. Choose any $\tilde\rho\in\mathbb{R}$, any compact subset $K\subset(0,\epsilon_3]$, and any interior point $\tilde\epsilon\in K$. By (3.24) and (3.25), the fact that for $(\rho,\epsilon)\in\mathbb{R}\times K$ the map $\mathscr{Q}_{\rho,\epsilon}$ is canonical and is defined on a ρ,ϵ-independent neighborhood of the origin (thereby having the intersection property in the second paragraph of Appendix A), the above analyticity properties of $\phi_{\rho,\epsilon}$, $\phi_{\rho,\epsilon}^{-1}$, $\lambda_{\rho,\epsilon}$, and $\mu(\rho,\epsilon)$, and the facts that at the latter ρ,ϵ values $|\lambda(\rho,\epsilon)| = 1$ and $\mu(\rho,\epsilon)$ is real and nonzero, we may apply Theorem B.1—the version of the twist theorem in Appendix B—to the map $\mathscr{Q}_{\rho,\epsilon}$. We thus conclude that at each of the latter ρ, ϵ there exists for each $\delta>0$ a simple closed curve $\Gamma_{\rho,\epsilon}$ of the form (B1) which is invariant under $\mathscr{P}_{\rho,\epsilon}$ and lies in the punctured disk $0<\|\zeta\|<\delta$. Hence, by (3.19) and the analyticity of $Z_1(\rho,\epsilon)$ in ρ, ϵ, $\Gamma_{\rho,\epsilon}+Z_1(\rho,\epsilon)$ is an invariant simple closed curve of $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ at each such ρ,ϵ which lies in the punctured disk $0<\|\zeta - Z_1(\rho,\epsilon)\|<\delta$ and is of the form $\{(x,y)\in\mathbb{R}^2: x = r(\varphi;\rho,\epsilon), y = s(\varphi;\rho,\epsilon)\}$, with r, s 2π-periodic in φ on $\mathbb{R}$ for $(\rho,\epsilon)\in\mathbb{R}\times K$ and analytic in φ, ρ, ϵ on $\mathbb{R}\times\mathbb{R}\times K$. *However, what is really important here is that r, s are jointly continuous in φ, ρ, ϵ on $\mathbb{R}\times\mathbb{R}\times K$ (see Ref. 10), rather than analytic.* An elementary argument which uses this joint continuity property entails that for $(\rho,\epsilon)\in\mathbb{R}\times K$ there exists a positive constant $d = d(\delta,\tilde\rho,\tilde\epsilon)$ such that

$$\|(\mathscr{P}_{\rm NR}(\rho,\epsilon))^n(\zeta)\|<\delta,\quad n = 1,2,...,$$

if

$$\|(\zeta,\rho,\epsilon) - (Z_1(\tilde\rho,\tilde\epsilon),\tilde\rho,\tilde\epsilon)\|<d$$

in terms of the usual $\mathbb{R}^4$ norm. By this stability result and a standard elementary argument of the same type as one used in Ref. 1, the orbital stability assertions of Theorem 2 follow for the case $H = H_{\rm NR}$. ■

IV. PROOF OF THE THEOREMS 1 AND 2 FOR THE R MODEL

For $E>0$, ϵ is defined in the present section by

$$\epsilon = E^{-1/2}. \tag{4.1}$$

This section has a structure parallel to that of Sec. III, its four subsections, Secs. IV A–IV D, being similar in content to Secs. III A–III D, respectively. The results of the present section can be proved by arguments of the same type as those invoked in Sec. III, and hence will be stated mostly without detailed proofs.

A. Auxiliary considerations

In this subsection, our sole assumptions on $V(x_1,x_2)$ are that it has property (I) and is one-periodic in x_1 in the strip (2.1).

For $H = H_{\rm NR}$, $E>0$, the solution $x_i^H(t;z,E)$, $p_i^H(t;z,E)$ $(i = 1,2)$, defined in Sec. II A, will be denoted by $x_i(t;z,\epsilon)$, $p_i(t;z,\epsilon)$ $(i = 1,2)$.

Instead of dealing with this solution directly, we will use the functions

$$X'(u;z,\epsilon) = x_2(t;z,\epsilon), \tag{4.2a}$$

$$P'(u;z,\epsilon) = p_2(t;z,\epsilon), \tag{4.2b}$$

where u is a new independent variable given by

$$u = x_1(t;z,\epsilon). \tag{4.2c}$$

By (2.6) and (4.2),

$$X'(\rho;z,\epsilon) = \xi, \tag{4.3a}$$

$$P'(\rho;z,\epsilon) = \eta. \tag{4.3b}$$

By (4.2) and the definition of $x_i(t;z,\epsilon)$, $p_i(t;z,\epsilon)$ stated in this subsection [recalling the condition $p_1(t;z,\epsilon)>0$], it follows that at those $z\in\mathbb{R}^3$, $\epsilon>0$ such that $(\rho,\xi)\in\mathscr{S}$ [see (2.1)] at which this solution exists, $X'(u;z,\epsilon)$, $P'(u;z,\epsilon)$ satisfy the isoenergetically reduced system

$$\frac{dX'}{du} = \epsilon^2\frac{P'}{\sigma_1(u,Z',\epsilon)}, \tag{4.4a}$$

$$\frac{dP'}{du} = \frac{[1-\epsilon^2V(u,X')]f(u,X')}{\sigma_1(u,Z',\epsilon)}, \tag{4.4b}$$

of Hamiltonian differential equations. Here $Z' = (X',P')$, f is defined by (3.5b) as before, and

$$\sigma_1(u,w,\epsilon) = \{[1-\epsilon^2V(u,w_1)]^2 - \epsilon^4(w_2^2+1)\}^{1/2}, \tag{4.5}$$

where $w = (w_1,w_2)$ as previously. For the denominators in (4.4), we have $\sigma_1(u;Z'(u;z,\epsilon),\epsilon) = \epsilon^2p_1(t;z,\epsilon)>0$ when $x_i(t;z,\epsilon)$, $p_i(t;z,\epsilon)$ $(i = 1,2)$ exist for given $t,z,\epsilon>0$ values.

The integral equations

$$X'(u;z,\epsilon) = \xi + \epsilon^2\int_\rho^u \frac{P'(u';z,\epsilon)}{\sigma_1(u',Z'(u';z,\epsilon),\epsilon)}\,du', \tag{4.6a}$$

$$P'(u;z,\epsilon) = \eta + \int_\rho^u \frac{[1-\epsilon^2V(u',X'(u';z,\epsilon))]f(u',X'(u';z,\epsilon))}{\sigma_1(u',Z'(u';z,\epsilon),\epsilon)}\,du', \tag{4.6b}$$

incorporating (4.3), will be as useful in the present section as Eqs. (3.6) were in the previous one.

As in Sec. III A, we will follow a procedure inverse to that in Ref. 1. Given that V has property (I), for each $z\in\mathbb{R}^3$, $\epsilon\geqslant 0$ with $(\rho,\xi)\in\mathscr{S}$ [see (2.1)] we *define* $X'(u;z,\epsilon)$, $P'(u;z,\epsilon)$ as a solution of (4.4) at all u in a maximum open interval in $\mathbb{R}$ containing $u = \rho$, which satisfies the initial conditions (4.3) and is such that $(u,X'(u;z,\epsilon))\in\mathscr{S}$ and $\sigma_1(u;Z'(u;z,\epsilon),\epsilon)>0$ at each u in this interval. This definition and the smoothness properties of the right sides of Eqs. (4.4) when V obeys (I) entail that this solution is unique. We will deduce the relevant properties of $x_i(t;z,\epsilon)$, $p_i(t;z,\epsilon)$ $(i = 1,2)$ from those of $X'(u;z,\epsilon)$, $P'(u;z,\epsilon)$.

The following lemma, analogous to Lemma 3.1 and proved similarly, holds.

Lemma 4.1: Let $V(x_1,x_2)$ obey (I) and be one-periodic in x_1 in the strip (2.1). Choose positive constants a,b as in Lemma 3.1. There exists a positive constant ϵ_0' such that $X'(u;z,\epsilon)$, $P'(u;z,\epsilon)$ exist for $u\in J'(\rho)$, $z = (\rho,\epsilon)\in\mathscr{Y}$ [see (3.7)], $0\leqslant\epsilon\leqslant\epsilon_0'$, $J'(\rho)$ being an ϵ-independent interval containing $[\rho,\rho+1]$ for all $\rho\in\mathbb{R}$. Moreover, at each such u,z,ϵ, the functions $X'(u;z,\epsilon)$ $P'(u;z,\epsilon)$ are real analytic in these variables and $(X'(u;z,\epsilon), P'(u;z,\epsilon))$ is in the region $\mathscr{U}_0$ defined in Lemma 3.1.

Next we define a mapping which is the inverse of that defined formally by (4.2c). At every $z\in\mathscr{Y}$, $0\leqslant\epsilon\leqslant\epsilon_0'$, let

$$u\mapsto t \tag{4.7a}$$

be the diffeomorphism from $J'(\rho)$ onto an interval in $\mathbb{R}$ via

$$t=\int_\rho^u \frac{[1-\epsilon^2 V(u',X(u';z,\epsilon)]}{\sigma_1(u',Z'(u';z,\epsilon),\epsilon)}du' = g_1(u;z,\epsilon)\ . \tag{4.7b}$$

Arguments similar to those used in Sec. III A in an analogous connection, including the use of Lemma 4.1, show that this is indeed a diffeomorphism. In fact, $g_1(u;z,\epsilon)$ is strictly increasing in u at the latter z,ϵ and real analytic in its arguments at the u,z,ϵ at which we defined it. We conclude that $g_1^{-1}(t;z,\epsilon)$ (g_1^{-1} = inverse of g_1) is analytic in its arguments for $(t,z,\epsilon)\in g_1(J'(\rho);z,\epsilon)\times\mathscr{Y}\times(0,\epsilon_0]$ $(g_1(J'(\rho);z,\epsilon)$ = image of $J'(\rho)$ under $u\mapsto t$ at given z,ϵ).

At each such $(z(t,\epsilon))$ we define $x_1(t;z,\epsilon)=g_1^{-1}(t;z,\epsilon)$, $p_1(t;z,\epsilon)=[E-V(x_1(t;z,\epsilon), x_2(t;z,\epsilon))]\dot{x}_1(t;z,\epsilon)$, and define $x_2(t;z,\epsilon), p_2(t;z,\epsilon)$ by (4.2a) and (4.2b). By the properties of g_1 and Lemma 4.1 in particular, one can readily show that these functions $x_i(t;z,\epsilon), p_i(t;z,\epsilon)$ ($i=1,2$) are analytic in their arguments at these t,z,ϵ values and agree there with the respective functions defined earlier in this subsection.

B. Definition of the map $\mathscr{P}_{\rm R}(\rho,\epsilon)$ and formulas for this map for $\epsilon\downarrow 0$

In this subsection, we will suppose that V satisfies (I) and (II).

Definition: For each $\rho\in\mathbb{R}$, $0\leqslant\epsilon\leqslant\epsilon_0'$, $\mathscr{P}_{\rm R}(\rho,\epsilon)$ is a mapping with domain $\mathscr{U}$ [see (3.7b)] which sends each $(\xi,\eta)\in\mathscr{U}$ into

$$(X'(\rho+1,z,\epsilon), P'(\rho+1,z,\epsilon))\in\mathbb{R}^2\ .$$

Remark: $\mathscr{P}_{\rm R}(\rho,\epsilon)$ is a well-defined canonical map and $X'(\rho+1,z,\epsilon)$, $P'(\rho+1,z,\epsilon)$ have the properties attributed to ξ', η' in the Remark after the definition of $\mathscr{P}_{\rm NR}(\rho,\epsilon)$ in Sec. III B, but with ϵ_0 replaced by ϵ_0'. This follows by the analyticity properties of $X'(u;z,\epsilon)$, $P'(u;z,\epsilon)$ stated in Lemma 4.1, an invariance property similar to one mentioned in the latter Remark, and the cited theorem of the book by Arnold and Avez.[14]

The next lemma is analogous to Lemma 3.2 and plays a comparable role in this section to that played by the latter lemma in Sec. III.

Lemma 4.2: For $z\in\mathscr{Y}$,

$$X'(\rho+1,z,\epsilon) = \xi+\epsilon^2\left[\eta+\int_\rho^{\rho+1}(\rho+1-u)f(u,\xi)du\right] + \epsilon^4\sum_{r=0}^{1}k_r(\rho,\xi)\eta^r+\epsilon^6\sum_{r=0}^{3}l_r(\rho,\xi)\eta^r+O(\epsilon^8)\ , \tag{4.8a}$$

$$P'(\rho+1;z,\epsilon) = \eta+\int_\rho^{\rho+1}f(u,\xi)du+\epsilon^2\left[\eta\int_\rho^{\rho+1}(u-\rho)f(u,\xi)du + \int_\rho^{\rho+1}du\,f(u,\xi)\int_\rho^u du'\,f(u',\xi)\right] + \epsilon^4\sum_{r=0}^{2}M_r(\rho,\xi)\eta^r+\epsilon^6\sum_{r=0}^{3}N_r(\rho,\xi)\eta^r+O(\epsilon^8)\ , \tag{4.8b}$$

where k_r, l_r, m_r, n_r are analytic in ρ,ξ for $\rho\in\mathbb{R}$, $|\xi|\leqslant a$ (a is as in Lemmas 3.1 and 4.1), are bounded there, and are η, ϵ-independent.

Proof: Analogous to that of Lemma 3.2. In particular, one expands the right- and left-hand sides of (4.7) in even powers of ϵ and equates coefficients. ■

Later on in this section, we will need the formulas

$$\frac{\partial X'(\rho+1;z,\epsilon)}{\partial\xi} = 1+\epsilon^2\int_\rho^{\rho+1}(\rho+1-u)f'(u,\xi)du+O(\epsilon^4)\ , \tag{4.9a}$$

$$\frac{\partial X'(\rho+1;z,\epsilon)}{\partial\eta} = \epsilon^2+O(\epsilon^4)\ , \tag{4.9b}$$

$$\frac{\partial P'(\rho+1;z,\epsilon)}{\partial\xi} = \int_0^1 f(u,\xi)du+O(\epsilon^2)\ , \tag{4.9c}$$

$$\frac{\partial P'(\rho+1;z,\epsilon)}{\partial\eta} = 1+\epsilon^2\int_\rho^{\rho+1}(u-\rho)f'(u,\xi)du+O(\epsilon^4)\ , \tag{4.9d}$$

which follow from (4.8) by arguments similar to those adduced to obtain (3.10) from (3.9).

C. Proof of Theorem 1 for $H=H_{\rm R}$

As in Sec. III C, we assume in the present subsection that V satisfies (I), (II), and (3.11).

The principal result of this subsection is Theorem 4.1, which together with simple arguments easily implies Theorem 1 for the case $H=H_{\rm R}$, except that the former theorem gives a much cruder estimate for $Z_H(\rho,\epsilon)$ for $\epsilon\downarrow 0$ in the case $H=H_{\rm NR}$ than does the latter. These arguments are analogous to those outlined for the NR model in the Remark to Theorem 3.1. Henceforth, we will write $Z_1'(\rho,\epsilon)$ for $Z_H(\rho,\epsilon)$ for $H=H_{\rm R}$. A proof that $Z_1'(\rho,\epsilon)$ is given by the estimates (2.8) appropriate to the R model will be sketched very briefly after providing the next theorem.

Theorem 4:1: Let V have properties (I), (II), and (3.11). Then:

(1′) There exist positive constants $\alpha'\leqslant\min\{a,b\}$, $\epsilon_1'\leqslant\epsilon_0'$ with a,b as in Lemma 3.1, such that for $0<\epsilon\leqslant\epsilon_1'$ Eqs. (4.4) have exactly one solution $X_1'(u,\epsilon)$, $P_1'(u,\epsilon)$ which exists for all $u\in\mathbb{R}$, is one-periodic in u, and for which $Z_1'(u,\epsilon) = (X_1'(u,\epsilon), P_1'(u,\epsilon))$ has norm $\|Z_1'(u,\epsilon)\|\leqslant\alpha'$ at each such u. [Note that $Z_1'(\rho,\epsilon)\in\mathscr{U}$ at these ρ, ϵ.]

(2′) The estimates

$$X_1'(u,\epsilon)=O(\epsilon)\ , \tag{4.10a}$$

$$P_1'(u,\epsilon)=O(1)\ , \tag{4.10b}$$

hold for $\epsilon\downarrow 0$, uniformly in u on $\mathbb{R}$.

(3′) For $A_2>0$ (resp. $A_2<0$) the ellipticity (resp. hyperbolicity) assertion (3) of Theorem 3.1 and the analyticity assertion (4) of that theorem hold with $X_1(u,\epsilon)$, $P_1(u,\epsilon)$, α, ϵ_1 replaced by $X'_1(u,\epsilon)$, $P'_1(u,\epsilon)$, α', ϵ'_1, respectively.

Proof: It consists of two major steps: (a′) proof of (1′) and (2′); (b′) proof of (3′).

(a′) In order to prove the existence and uniqueness of the periodic solution in the theorem, we make the noncanonical transformation of coordinates

$$X^*=X', \quad P^*=\epsilon P' \tag{4.11}$$

for $\epsilon>0$, where the recall that for $E>0$, ϵ is defined by (4.1) of the present section. In terms of the new coordinates, Eqs. (4.4) take the form

$$\frac{dZ^*}{du}=\epsilon\phi^*(u,Z^*,\epsilon)\,, \tag{4.12a}$$

where $Z^*=(X^*,P^*)$ and

$$\phi^*(u,w,\epsilon)=\{w_2/\sigma_1^*(u,w,\epsilon)\,,\; \times[1-\epsilon^2V(u,w_1)]\,f(u,w_1)/\sigma_1^*(u,w,\epsilon)\}\,, \tag{4.12b}$$

with $w=(w_1,w_2)$ and

$$\sigma_1^*(u,w,\epsilon)=\{[1-\epsilon^2V(u,w_1)]^2-\epsilon^2(w_2^2+\epsilon^2)\}^{1/2}. \tag{4.12c}$$

Therefore

$$\int_0^1\phi^*(u,0,0)du=0\,, \tag{4.13a}$$

$$\int_0^1\frac{\partial\phi^*(u,0,0)}{\partial w}\,du=N\,, \tag{4.13b}$$

by (2.2), (2.3), (3.5b), (4.12b), and (4.12c), where N is as in (3.14b). Equations (4.11)–(4.13) and the facts that N is a nonsingular matrix and that $\phi^*(u,w,\epsilon)$ is one-periodic in u and analytic in u,w,ϵ in a neighborhood of the subset $\{(u,w=0,\epsilon=0),\ u\in\mathbb{R}\}\subset\mathbb{R}^4$ allow us to apply Theorem A.1 of Appendix A and the remark thereto to the transformed system (4.12a) and hence to conclude that assertions (1′) and (2′) of the theorem hold for the original system (4.4).

(b′) Let $M'_1(\rho,\epsilon)$ denote the Jacobian $\partial\mathscr{P}_{\rm R}(\rho,\epsilon)(\zeta)/\partial\zeta$ of the map $\mathscr{P}_{\rm R}(\rho,\epsilon)$ at its fixed point $\zeta=Z'_1(\rho,\epsilon)$. By the definition of $\mathscr{P}_{\rm R}(\rho,\epsilon)$ and by (4.9a), (4.9d), (3.15), and (4.10a),

$$\operatorname{Tr}M'_1(\rho,\epsilon)=2-A_2\epsilon^2+O(\epsilon^4) \tag{4.14}$$

for $\rho\in\mathbb{R}$. By (4.14) and since the equation $\det M'_1(\rho,\epsilon)=1$ holds at each such ρ for $0<\epsilon\leqslant\epsilon'_1$, where $\epsilon'_1\leqslant\epsilon'_0$ is a sufficiently small positive constant, it follows by arguments virtually identical to the ones in part (b) of the proof of Theorem 3.1 that $Z'_1(\rho,\epsilon)$ is an elliptic (resp. ordinary hyperbolic) fixed point at each such ρ,ϵ, if $A_2>0$ (resp. $A_2<0$).

(c′) The analyticity assertion in (3′) follows by using (4.10b), (4.11), and Theorem A2. ■

A systematic procedure for determining the coefficients of the series for $X'_1(\rho,\epsilon)$, $P'_1(\rho,\epsilon)$ in powers of ϵ is available, but we will confine ourselves here to proving that these two functions are given by the right-hand sides of (2.8a) and (2.8b), respectively, appropriate to the R model. By using the analog of (3.17b) for this model at $\zeta=Z'_1(\rho,\epsilon)$, together with (II), (2.4a), (3.5b), (4.8b), and (4.10), we find

$$X'_1(\rho,\epsilon)=O(\epsilon^2)\,. \tag{4.15a}$$

By (4.8a), (4.15a), and arguments analogous to those adduced to obtain (4.15b), it follows that

$$P'_1(\rho,\epsilon)=h(\rho)+O(\epsilon^2)\,. \tag{4.15b}$$

Hence we have derived the promised result for $Z'_1(\rho,\epsilon)$.

D. Proof of Theorem 2 for $H=H_{\rm R}$

In this subsection, V has properties (I)–(III).

Consider $\rho\in\mathbb{R}$, $0\leqslant\epsilon\leqslant\epsilon'_1$. We define $\mathscr{P}'_{\rho,\epsilon}$ by

$$\mathscr{P}'_{\rho,\epsilon}(\zeta)=\mathscr{T}'^{-1}_{\rho,\epsilon}\circ\mathscr{P}_{\rm R}(\rho,\epsilon)\circ\mathscr{T}'_{\rho,\epsilon}(\zeta)\,, \tag{4.16}$$

for $\zeta\in\mathscr{U}-Z'_1(\rho,\epsilon)$, $\mathscr{T}'_{\rho,\epsilon}$ being the translation $\zeta\mapsto\zeta-Z'_1(\rho,\epsilon)$. Since $\zeta=Z'_1(\rho,\epsilon)$ is a fixed point of $\mathscr{P}_{\rm R}(\rho,\epsilon)$, $\zeta=0$ is one of $\mathscr{P}'_{\rho,\epsilon}$. Letting $(\hat{\xi},\hat{\eta})=\mathscr{P}'_{\rho,\epsilon}(\xi,\eta)$ the series

$$\hat{\xi}=\sum_{n=0}^{\infty}\sum_{0\leqslant s\leqslant n}[(n-s)!s!]^{-1}\left(\frac{\partial^n\tilde{\xi}}{\partial\xi^{n-s}\partial\eta^s}\right)_1\xi^{n-s}\eta^s\,, \tag{4.17a}$$

$$\hat{\eta}=\sum_{n=0}^{\infty}\sum_{0\leqslant s\leqslant n}[(n-s)!s!]^{-1}\left(\frac{\partial^n\tilde{\eta}}{\partial\xi^{n-s}\partial\eta^s}\right)_1\xi^{n-s}\eta^s\,, \tag{4.17b}$$

converge at the latter ρ,ϵ if (ξ,η) is in a ρ,ϵ-independent rectangle centered at the origin, where $\tilde{\xi}=X'(\rho+1;z,\epsilon)$, $\tilde{\eta}=P'(\rho+1;z,\epsilon)$, and where the subscript 1 signifies that the derivatives are evaluated at $\zeta=Z'_1(\rho,\epsilon)$. Reasons analogous to those stated to prove a similar property of the series (3.20) ensure the existence of such a rectangle.

In order to prove Theorem 2 for $H=H_{\rm R}$, we will reduce $\mathscr{P}'_{\rho,\epsilon}$ to Birkhoff normal form, using the results of the next lemma.

Lemma 4.3: For $\rho\in\mathbb{R}$,

$$\left(\frac{\partial\tilde{\xi}}{\partial\xi}\right)_1=1+\epsilon^2\int_\rho^{\rho+1}(1+\rho-u)f'(u,0)du+O(\epsilon^4)\,, \tag{4.18a}$$

$$\left(\frac{\partial\tilde{\xi}}{\partial\eta}\right)_1=\epsilon^2+O(\epsilon^4)\,, \tag{4.18b}$$

$$\left(\frac{\partial\tilde{\eta}}{\partial\xi}\right)_1=\int_0^1 f'(u,0)du+O(\epsilon^2)\,, \tag{4.18c}$$

$$\left(\frac{\partial\tilde{\eta}}{\partial\eta}\right)_1=1+\epsilon^2\int_\rho^{1+\rho}(u-p)f'(u,0)du+O(\epsilon^4)\,, \tag{4.18d}$$

and for $n=2,3$,

$$\left(\frac{\partial^n\tilde{\xi}}{\partial\xi^{n-s}\partial\eta^s}\right)_1=O(\epsilon^{2+s}),\quad 0\leqslant s\leqslant 2\,, \tag{4.19a}$$

$$\left(\frac{\partial^3\tilde{\xi}}{\partial\eta^3}\right)_1=O(\epsilon^6)\,, \tag{4.19b}$$

$$\left(\frac{\partial^n\tilde{\eta}}{\partial\xi^{n-s}\partial\eta^s}\right)_1=O(\epsilon^{2s}),\quad 1\leqslant s\leqslant n\,. \tag{4.19c}$$

Proof of Lemma 4.3: Equations (4.18) are an immediate consequence of (4.9), (3.15), (4.15), and the one-periodicity of $V(x_1,x_2)$ in x_1. The proof of (4.19) is similar. ■

Let $\epsilon_2' \leqslant \epsilon_1'$ be a small enough positive constant. Then for $\rho \in \mathbb{R}$, $0 < \epsilon \leqslant \epsilon_2'$, the appropriately labeled eigenvalues of the Jacobian $M_1'(\rho,\epsilon)$ of the map $\mathscr{P}_{\rm R}(\rho,\epsilon)$ at $\zeta = Z_1'(\rho,\epsilon)$ are of the form $\lambda'(\rho,\epsilon)$, $\overline{\lambda'(\rho,\epsilon)}$ and have the remaining properties (i) and also the properties (ii) possessed by the eigenvalues of $M_1(\rho,\epsilon)$ at $\zeta = Z_1(\rho,\epsilon)$ (see Sec. III D). This follows by arguments similar to those used in the case $H = H_{\rm NR}$.

Let $\rho \in \mathbb{R}$, $0 < \epsilon \leqslant \epsilon_2'$. Then the general theory of reduction to Birkhoff normal form,[16,17] the fact that $\mathscr{P}'_{\rho,\epsilon}$ is canonical, that the series (4.17) converge in a ρ,ϵ-independent neighborhood of the origin, the above properties of $\lambda'(\rho,\epsilon)$, (4.18) and (4.19) and other arguments, analogous to those used in the reduction of $\mathscr{P}_{\rho,\epsilon}$ entail that $\mathscr{P}'_{\rho,\epsilon}$ is conjugate to a canonical mapping $\mathscr{Q}'_{\rho,\epsilon}$ in this form:

$$\mathscr{Q}'_{\rho,\epsilon} = \phi'^{-1}_{\rho,\epsilon} \circ \mathscr{P}'_{\rho,\epsilon} \circ \phi'_{\rho,\epsilon} | B'_{\rho,\epsilon} ,$$

where $\phi'_{\rho,\epsilon}$ is a diffeomorphism from the open ball $B'_{\rho,\epsilon} = \{(x,y) \in \mathbb{R}^2 : \|(x,y)\| < r'_{\rho,\epsilon}\}$, $r'_{\rho,\epsilon} > 0$, onto an open neighborhood of $0 \in \mathbb{R}^2$ and is one-periodic in ρ. The second and third sentences after (3.24) apply to $\phi'_{\rho,\epsilon}$ if $B_{\rho,\epsilon}$, $r_{\rho,\epsilon}$, $\phi_{\rho,\epsilon}$, $\phi^{-1}_{\rho,\epsilon}$, ϕ_i, ψ_i, K are replaced by corresponding primed symbols. Naturally, reasons analogous to those mentioned in the paragraph containing (3.24) entail that for each $\rho \in \mathbb{R}$ and ϵ in a compact subset of $(0,\epsilon_2']$, $\mathscr{Q}'_{\rho,\epsilon}$ is defined on a ρ, ϵ-independent neighborhood of the origin. Writing

$$(x_1',y_1') = \mathscr{Q}'_{\rho,\epsilon}(x,y) ,$$

$$z = x + iy, \quad z_1' = x_1' + iy_1' ,$$

we have

$$z_1' = \lambda'(\rho,\epsilon) z[1 + \mu'(\rho,\epsilon)|z|^2] + O(|z|^4), \quad z \to 0 ,$$

where for $\rho \in \mathbb{R}$, $0 < \epsilon \leqslant \epsilon_2'$ the first twist coefficient $\mu'(\rho,\epsilon)$ is real, analytic in ρ,ϵ and one-periodic in ρ, and

$$\mu'(\rho,\epsilon) = \tfrac{1}{8} A_2^{-2}(A_2 A_4 - \tfrac{5}{3} A_3^2)\epsilon^2 + O(\epsilon^3) , \qquad (4.20)$$

uniformly in ρ on $\mathbb{R}$. This equation was derived by choosing the matrix $C'(\rho,\epsilon)$ diagonalizing $M_1'(\rho,\epsilon)$, i.e., satisfying Eq. (3.26) with appropriate primed symbols, to be normalized in the same way as $C(\rho,\epsilon)$ and to be analytic in ρ,ϵ and one periodic for $\rho \in \mathbb{R}$, $0 < \epsilon \leqslant \epsilon_2'$. Except for the uniformity property, which follows similarly to how the corresponding property of (3.27) was proved, (4.20) is the same as the equation obtained in Ref. 1 (see Lemma 4.5 therein) for the first twist coefficient pertaining to a map [called $\mathscr{P}_{\rm R}(E)$ there], which is $\mathscr{P}'_{\rho,\epsilon}$ specialized to the situation when $Z_1'(\rho,\epsilon) \equiv 0$. Analogously to what was said respecting a similar property of the first twist coefficient of $\mathscr{P}_{\rho,\epsilon}$ this can be seen by (1) the facts that (4.18) and (4.19) also hold for the respective derivatives evaluated at $\zeta = 0$, and the way in which $\lambda'(\rho,\epsilon)$ and $C'(\rho,\epsilon)$ were chosen; and (2) an appropriate version of Eq. (3.30) on Ref. 1.

Proof of Theorem 2 for $H = H_{\rm R}$: This result now follows by arguments that are verbatim repetitions of those marshalled to prove this theorem for $H = H_{\rm NR}$ in Sec. III D. Equation (4.20) and other results in the last paragraph play an important role in the proof of this theorem for the R model. ■

ACKNOWLEDGMENT

We wish to thank J. A. Ellison for many valuable conversations, and particularly for one which motivated the simple power-series proofs of Lemmas 3.2 and 4.2.

APPENDIX A: EXISTENCE, UNIQUENESS, ASYMPTOTIC BEHAVIOR, AND ANALYTICITY OF PERIODIC SOLUTIONS OF CERTAIN ODE SYSTEMS WITH PERIODIC COEFFICIENTS NEAR SINGULAR POINTS OF THE CORRESPONDING AVERAGED SYSTEMS

To make this paper as self-contained as possible, we derive in this appendix a version of certain results on periodic solutions proved previously by Hale[6] and Swinnington-Dyer[7] by using averaging-theory methods.[18] Our derivation is simpler than those in Refs. 6 and 7 and presupposes no knowledge of averaging theory. In this appendix, we also prove Theorem A.2., which deals with the analyticity of the periodic solutions of Theorem A.1 under suitable hypotheses and is similar to a result established by Swinnerton-Dyer[19] by a different method. Theorems A.1 and A.2 are used in Secs. III and IV to prove certain key properties of the periodic solutions mentioned in Theorem 1 of Sec. II A.

Consider the nth-order system

$$\frac{dw}{du} = \epsilon F(u,w,\epsilon) , \qquad (A1)$$

where F is a function from $\mathbb{R} \times U_0 \times [0,\epsilon_0]$ into $\mathbb{R}^2$, with $U_0 = \{x \in \mathbb{R}^n : |x_i| \leqslant \alpha_0,\ i = 1,...,n\}$, α_0, and ϵ_0 being positive constants. We will assume: (A) F is one-periodic in u, (B) F is continuous, (C) $\partial F/\partial w$ exists as a continuous function from $\mathbb{R} \times U_0 \times [0,\epsilon_0] \to \mathbb{R}^n$, and (D) one has

$$\int_0^1 F(u,0,0)\, du = 0 , \qquad (A2)$$

and the constant matrix

$$A = \int_0^1 \frac{\partial F(u,0,0)}{\partial w}\, du \qquad (A3)$$

is nonsingular.

Theorem A.1: Let F be as stated in the last paragraph. Then there exist positive constants $\epsilon^* \leqslant \epsilon_0$ and $\sigma < \alpha_0$ such that for $u \in \mathbb{R}$, $0 < \epsilon \leqslant \epsilon^*$ system (A1) has exactly one solution $w^*(u,\epsilon)$ in $B(\sigma) = \{x \in \mathbb{R}^n : \|x\| \leqslant \sigma\}$ which is one-periodic in u. For $\epsilon \downarrow 0$, $w^*(u,\epsilon) \to 0$, uniformly in u on $\mathbb{R}$. Moreover, if in addition

$$\int_0^1 F(u,0,\epsilon) du = O(\epsilon) , \qquad (A4)$$

then for $\epsilon \downarrow 0$

$$\|w^*(u,\epsilon)\| = O(\epsilon) \qquad (A5)$$

holds in this limit in this uniform sense.

Remark: The fact that (A4) holds in this sense under the relatively weak conditions on F in Theorem A.1 is of mathematical interest and is proved below. However, this proof is unnecessary for the needs of the present paper. Indeed, in the applications in Secs. III and IV, the much stronger conditions of Theorem A.2 are satisfied. Under those conditions, the uniform estimate (A5) follows immediately from the results: (a) by the latter theorem, $w^*(u,\epsilon)$ is one-periodic in u and analytic in its arguments on $\mathbb{R}\times[0,\tilde{\epsilon}]$ for some small enough positive constant $\tilde{\epsilon}$; (b) $w^*(u,0)=0$ by results in Part (1) of the proof of Theorem A.1.

Proof of Theorem A.1: It consists of two main steps: (1) proof of existence and uniqueness of $w^*(u,\epsilon)$; (2) proof of (A4).

(1) Write $z=(\rho,\zeta)$ for $(\rho,\zeta)\in\mathbb{R}\times\mathbb{R}^n$ and let $V_0=\mathbb{R}\times U_0$. At each $(z,\epsilon)\in V_0\times[0,\epsilon_0]$, let $w(u;z,\epsilon)$ be the unique solution of (A1) satisfying the initial condition

$$w(\rho;z,\epsilon)=\zeta \tag{A6}$$

and existing at each u is a maximal open interval in $\mathbb{R}$ containing $u=\rho$.

Applying the usual successive-approximation arguments to the integral equation

$$w(u;z,\epsilon)=\zeta+\epsilon\int_\rho^u F(u',w(u';z,\epsilon),\epsilon)du' \tag{A7}$$

one sees that $w(u;z,\epsilon)$ exists for

$$u\in I(\rho),\quad z\in V,\quad 0\leqslant\epsilon\leqslant\epsilon_0, \tag{A8}$$

where $V=\mathbb{R}\times U$, $U=\{x\in\mathbb{R}^n: |x_i|\leqslant\alpha,\ i=1,\ldots,n\}$, and ϵ_0 is a positive constant so small that $I(\rho)=[\rho-(\alpha_0-\alpha)/\epsilon_0M,\ \rho+(\alpha_0-\alpha)/\epsilon_0M]$ contains the interval $[\rho,\rho+1]$ for $\rho\in\mathbb{R}$, α being a positive constant less than α_0 and $M=\max\|F(u,w,\epsilon)\|$ over $\mathbb{R}\times U_0\times[0,\epsilon_0]$. Trivially,

$$w(u;z_0,0)=0 \tag{A9}$$

for all $u\in\mathbb{R}$, where $z_0=(\rho,\zeta=0)$. Moreover, at each u,z,ϵ satisfying (A8), $w(u;z,\epsilon)\in U$, and $w(u;z,\epsilon)$ and $\partial w(u;z,\epsilon)/\partial\zeta$ are jointly continuous functions of their arguments.

At each $(\rho,\epsilon)\in\mathbb{R}\times[0,\epsilon_0]$ we define a nonlinear map $\mathscr{P}(\rho,\epsilon)$ from U into $\mathbb{R}^n$:

$$\mathscr{P}(\rho,\epsilon)(\zeta)=w(\rho+1;z,\epsilon). \tag{A10a}$$

By (A7) and (A10a), this map can be expressed in the form

$$\mathscr{P}(\rho,\epsilon)(\zeta)=\zeta+\epsilon[A\zeta+K(z,\epsilon)], \tag{A10b}$$

where

$$K(z,\epsilon)=\int_\rho^{\rho+1}\left[F(u,w(u;z,\epsilon),\epsilon)-\frac{\partial F(u,0,0)}{\partial w}\zeta\right]du. \tag{A10c}$$

By (A10b) and the invertibility of A, $\zeta\in U$ is a fixed point of the map $\mathscr{P}(\rho,\epsilon)$ for given $\rho\in\mathbb{R}$, $0<\epsilon\leqslant\epsilon_0$ iff

$$\zeta=-A^{-1}K(z,\epsilon). \tag{A11}$$

Notice that this statement is generally false for $\epsilon=0$.

The functions $K(z,\epsilon)$ and $\partial K(z,\epsilon)/\partial\zeta$ have the properties: (a) they vanish at $(z,\epsilon)=(z_0,0)$ for all $\rho\in\mathbb{R}$; (b) they are uniformly continuous functions of their arguments at each point (z,ϵ) of the closed set $V\times[0,\epsilon_0]$. Property (a) follows by (A2), (A9), and (A10c). Property (b) follows by (A7), (A9), (A10c), $I(\rho)\supset[\rho,\rho+1]$, the continuity and differentiability attributes of F and w, and the one-periodicity of $K(z,\epsilon)$ and $\partial K(z,\epsilon)/\partial\zeta$ in ρ. This periodicity is entailed by (A10c) and the equation

$$w(u;(\rho,\zeta),\epsilon)=w(u+1;(\rho+1;\zeta),\epsilon), \tag{A12}$$

which obtains for values of u,z,ϵ fulfilling (A8) and which can be derived by using, in particular, the one-periodicity of F in u and a standard uniqueness argument.

By (a) and (b), it follows easily by mimicking a proof[20] of the usual C^1 version of the implicit function theorem that there exist positive constants $\epsilon^*\leqslant\epsilon_0$ and σ, such that $B(\sigma)\subset U$ and that for all $\rho\in\mathbb{R}$, $0\leqslant\epsilon\leqslant\epsilon^*$ the operator $T(\rho,\epsilon)$ and $B(\sigma)$ defined by

$$T(\rho,\epsilon)(\zeta)=-A^{-1}K(z,\epsilon) \tag{A13}$$

maps $B(\sigma)$ into $B(\sigma)$, is continuous in ρ,ϵ, and is contractive, with contraction constant $k\in[0,1)$ independent of ρ,ϵ. Moreover, $T(\rho,0)$ has the fixed point $\zeta=0$ at each such ρ. Hence at every such ρ,ϵ, (i) Eq. (A11) has one and only one solution $\zeta=w^*(\rho,\epsilon)$ in $B(\sigma)$, i.e., $w^*(\rho,\epsilon)$ is the only fixed point of $\mathscr{P}(\rho,\epsilon)$ if in addition $\epsilon>0$; and (ii) $w^*(\rho,\epsilon)$ is continuous in ρ,ϵ (Ref. 21) and is therefore $o(1)$ for $\epsilon\downarrow0$ in the stated uniform sense. Standard arguments now show that, at each $0<\epsilon\leqslant\epsilon^*$, $w^*(u,\epsilon)$ is the unique periodic solution mentioned in the theorem. In particular, the one-periodicity of $w^*(u,\epsilon)$ at each such ϵ follows by straightforward arguments based on (A12) and the uniqueness of solution of (A11) in $B(\sigma)$.

(2) We now prove (A5). By (A13) and a well-known contraction-mapping estimate,[22]

$$\|w^*(\rho,\epsilon)\|\leqslant\|T(\rho,\epsilon)(0)\|/(1-k)$$
$$\leqslant\|A^{-1}\|/(1-k)\|K(z_0,\epsilon)\|,$$

for $\rho\in\mathbb{R}$, $0\leqslant\epsilon\leqslant\epsilon^*$. In particular, (A), (C), (A4), (A5), (A7), (A10c), the boundedness of $F(u,w,\epsilon)$ on $\mathbb{R}\times U_0\times[0,\epsilon_0]$ and the fact that $w(u;z_0,\epsilon)\in U\subset U_0$ for $u\in[\rho,\rho+1]$, $0\leqslant\epsilon\leqslant\epsilon^*\leqslant\epsilon_0$, we find

$$\begin{aligned}\|K(z_0,\epsilon)\|&\leqslant\int_\rho^{\rho+1}\|F(u,w(u;z_0,\epsilon),\epsilon)-F(u,0,\epsilon)\|du\\&\quad+\left\|\int_\rho^{\rho+1}F(u,0,\epsilon)du\right\|\\&\leqslant L\int_\rho^{\rho+1}\|w(u;z_0,\epsilon)\|du+K\epsilon\\&\leqslant L\epsilon\int_\rho^{\rho+1}du\int_\rho^u du'\|F(u',w(u';z_0,\epsilon),\epsilon)\|+K\epsilon\\&\leqslant\tfrac{1}{2}LM\epsilon+K\epsilon,\end{aligned}$$

where K and L are positive constants. Hence (A5) holds in the desired uniform sense under the stated hypotheses. ■

Theorem A.2: Let F be as in the second paragraph of this Appendix. Moreover, let $F(u,w,\epsilon)$ be real analytic in u,w,ϵ for $((u,w),\epsilon)\in V_0\times[0,\epsilon^{**}]$, where $\epsilon^{**}\leqslant\epsilon^*\leqslant\epsilon_0$ is a positive constant. Then the periodic solution $w^*(u,\epsilon)$ of (A1) defined on $\mathbb{R}\times[0,\epsilon^*]$ in the proof of Theorem A.1 is analytic in its arguments for $(u,\epsilon)\in\mathbb{R}\times[0,\epsilon^{***}]$ for some positive constant $\epsilon^{***}\leqslant\epsilon^{**}$.

Proof: By the proof of Theorem A.1, we know that the function g: $U\times\mathbb{R}\times[0,\epsilon_0]\to\mathbb{R}^n$ acting by

$$g(\zeta,\rho,\epsilon)=A\zeta+K(z,\epsilon) \tag{A14}$$

vanishes at the unique point $\zeta = w^*(\rho,\epsilon)$ in $B(\sigma) = \{x\in\mathbb{R}^n: \|x\|\leqslant\sigma\}\subset U_0$. In addition, there exists a positive constant $\hat{\epsilon}\leqslant\epsilon^{**}$ such that at each $\rho\in\mathbb{R}$, $0\leqslant\epsilon\leqslant\hat{\epsilon}$ the Jacobian matrix of the transformation $\zeta\mapsto g(\zeta,\rho,\epsilon)$ at $\zeta = w^*(\rho,\epsilon)$ is nonsingular. Indeed,

$$\frac{\partial g(w^*(\rho,\epsilon),\rho,\epsilon)}{\partial\zeta} = A + \frac{\partial K((\rho,w^*(\rho,\epsilon)),\epsilon)}{\partial\zeta} = A + o(1)$$

for $\epsilon\downarrow 0$, uniformly in ρ, where we have used properties (a) and (b) of $K(z,\epsilon)$ and $\partial K(z,\epsilon)/\partial\zeta$ [see the paragraph after (A11)], and the uniform estimate $\|w^*(\rho,\epsilon)\| = o(1)$ for $\epsilon\downarrow 0$.

Furthermore, if $u\in[\rho,\rho+1]$, $z\in V$, $0\leqslant\epsilon\leqslant\epsilon^{***}$, where $\epsilon^{***}\leqslant\hat{\epsilon}$ is a sufficiently small positive constant, then under the present hypotheses on F, $w(u;z,\epsilon)\subset U_0$, as we know from the proof of Theorem A.1, and $w(u;z,\epsilon)$ is real analytic in u,z,ϵ, as follows by standard arguments of the type used to prove Lemma 3.1. By these results, together with (A10c) and (A14), we see that $g(\zeta,\rho,\epsilon)$ is real analytic in z,ϵ at the latter z,ϵ values under these hypotheses.

By the properties of $g(\zeta,\rho,\epsilon)$ mentioned in the last two paragraphs and the usual analytic version of the implicit function theorem, it follows that for $\rho\in\mathbb{R}$, $0\leqslant\epsilon\leqslant\epsilon^{***}$ the unique solution $w^*(\rho,\epsilon)$ of $g(\zeta,\rho,\epsilon) = 0$ in $B(\sigma)$ is analytic in ρ,ϵ at such ρ,ϵ. ■

APPENDIX B: EXISTENCE AND ANALYTICITY PROPERTIES OF INVARIANT CURVES OR ANALYTIC AREA-PRESERVING MAPS IN THE PLANE

In this appendix we will state a slight generalization of the version of Moser's twist theorem stated in the Appendix of Ref. 1. This generalized result will be used in the proof of Theorem 2 of Sec. II A.

Let $\{M_\nu, \nu\in P\}$ be a family of mappings, where $P\subset\mathbb{R}^n$ is a compact set of the form $\{(\nu_1,\ldots,\nu_n)\in\mathbb{R}^n: \alpha_i\leqslant\nu_i\leqslant\beta_i, i = 1,\ldots n\}$. Each M_ν of this family maps every (x,y) lying in a ν-independent neighborhood $W\subset\mathbb{R}^2$ of the origin into $(x_1,y_1)\in\mathbb{R}^2$ by

$$z_1 = \gamma_1(\nu)z[1 + i\gamma_2(\nu)|z|^2] + S(z,\bar{z},\nu)\,,$$

where $z = x + iy$, $z_1 = x_1 + iy_1$. For $\nu = (\nu_1,\ldots\nu_n)$, γ_1 and γ_2 are analytic functions of ν from P into $\mathbb{C}$ and $\mathbb{R}$, respectively, such that

$$|\gamma_1(\nu)| = 1, \quad \gamma_2(\nu)\neq 0\,,$$

and S is analytic in $z,\bar{z},\nu$ for $((x,y),\nu)\in W\times P$ and such that

$$S(z,\bar{z},\nu) = O(|z|^4),\ z\to 0$$

at each such $\nu\in P$. Moreover, at all such ν, every circle in W centered at the origin intersects its image under M_ν.

Theorem B.1 (Moser's twist theorem): Under the hypotheses on the family $\{M_\nu, \nu\in P\}$ in the preceding paragraph, there exists for each $\epsilon>0$ and $\nu\in P$ an invariant closed curve of M_ν of the form

$$x = p(\xi,\nu), \quad y = q(\xi,\nu), \quad \xi\in\mathbb{R}\,, \tag{B1}$$

lying in the intersection of W with the punctured disk $0<x^2+y^2<\epsilon^2$ in the plane. Here, p, q are real analytic functions of ξ, ν for $(\xi,\nu)\in\mathbb{R}\times P$ which are 2π-periodic in ξ over $\mathbb{R}$ at each $\nu\in P$.

Proof: Similar to that of Theorem A.1 of Ref. 1. ■

[1]A. W. Sáenz, J. Math. Phys. **27**, 1925 (1985). The following errata occur in that paper: (a) in p. 1929, left column, change $X(u/\zeta,E)$ in Eq. (3.12a) to $X(u;\zeta,E)$, $f_2(u;\zeta,E)$ in Eq. (3.12b) to $f_2(u,X(u;\zeta,E))$, and $f_2(u';\zeta,E)$ in Eq. (3.13b) to $f_2(u',X'(u';\zeta,E))$; (b) in p. 1933, left column, change $V(u';X'(u';\zeta,E))$ to $V(u',X'(u',\zeta,E))$.

[2]A summary of this work is given in my paper on "Rigorous stability results on crystal channeling via canonical maps," in *Local and Global Methods of Nonlinear Dynamics*, edited by A. W. Sáenz, W. W. Zachary, and R. Cawley, *Lecture Notes in Physics*, Vol. 252 (Springer, New York, 1986), p. 231.

[3]D. S. Gemell, Rev. Mod. Phys. **46**, 129 (1974).

[4]C. Lehmann, *Interaction of Radiation with Solids and Elementary Defect Production* (North-Holland, Amsterdam, 1977).

[5]In this paper, $\mathbb{R}^n$ is regarded as a vector space whose elements are *column* vectors $(x_1,\ldots x_n)$ with n real components, the row-vector notation being used for typographical convenience. The usual Euclidean norm of $\mathbb{R}^n$ $(n\geqslant 2)$ will be denoted by $\|\cdot\|$.

[6]J. K. Hale, *Ordinary Differential Equations* (Krieger, Huntington, NY, 1980), 2nd ed., Theorem 3.2, p. 194.

[7]P. Swinnginton-Dyer, Proc. London Math. Soc. (3) **34**, 385 (1977), Theorem 3 and its corollary.

[8]C. L. Siegel and J. K. Moser, *Lectures on Celestial Mechanics* (Springer, Berlin, 1971), p. 228.

[9]See, e.g., H. S. Dumas and J. A. Ellison, "Particle channeling in crystals and the method of averaging," in *Local and Global Methods of Nonlinear Dynamics*, edited by A. W. Sáenz, W. W. Zachary, and R. Cawley, *Lecture Notes in Physics*, Vol. 252 (Springer, New York, 1986), p. 200.

[10]All statements in this paper asserting the analyticity of a real or complex-valued function in a certain set of scalar and/or vector arguments should be understood as joint analyticity in all the scalar arguments and all components of the vector arguments. Analogously, continuity in several variables will always be understood as joint continuity.

[11]See, e.g., J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer, New York, 1983), Theorem 1.4.1, p. 18.

[12]For the general theory, see, e.g., E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Dover, New York, 1944), 4th ed., Chap. XII, Sec. 141; A Wintner, *The Analytical Foundations of Celestial Mechanics* (Princeton U.P., Princeton, NJ, 1941), especially Secs. 180–182.

[13]The canonical (= symplectic) maps considered here are two-dimensional local versions of those defined, e.g., in V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, Berlin, 1978), p. 239. Such canonical maps in the plane preserve *oriented* areas.

[14]See, e.g., V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968), Theorem A.31.2, p. 231.

[15]Replacing (3.18) by the rough estimate (3.12) in this proof leads to formulas for the derivatives (3.21) and (3.22) which, although less precise in some cases than those given in Lemma 3.3, suffice in the proof of Theorem 2 for $H = H_{\rm NR}$. A similar remark applies to the replacement of (4.15) by the cruder estimates (4.10) in the proof of Lemma 4.3.

[16]W. Klingenberg, *Lectures on Closed Geodesics* (Springer, Berlin, 1977), pp. 100–103.

[17]See Ref. 8, Sec. 23.

[18]Theorem 3.1 of J. Murdock and C. Robinson, J. Differ. Eqs. **36**, 425 (1980), and averaging-theory arguments yield versions of Theorem 3.2 in p. 194 of Ref. 6 and Theorem 7 of Ref. 7. See also Theorem 5.4 of R. C. Churchill, M. Kummer, and D. L. Rod, J. Differ. Eqs. **49**, 359 (1983).

[19]See Ref. 7, Theorem 4.

[20]See Ref. 6, pp. 8,9.

[21]See, e.g., Ref. 6, Theorem 3.2, p. 7.

[22]See, e.g., V. Hutson and J. S. Pym, *Applications of Functional Analysis and Operator Theory* (Academic, London, 1980), Theorem 4.3.4, p. 116.

A formula on the Wiener–Hermite expansion

J. Nakayama
Department of Electronics, Kyoto Institute of Technology, Matsugasaki, Kyoto 606, Japan

(Received 13 April 1987; accepted for publication 1 July 1987)

Two mathematical formulas are given in an explicit form: one develops a product of two different multiple Wiener integrals in a series of Wiener–Hermite expansions; the other one develops a product of two random variables, each of which is described by a Wiener–Hermite expansion, in a series of another Wiener–Hermite expansion.

I. INTRODUCTION

As is well known, the Wiener–Hermite expansion technique[1–5] makes it possible to expand a nonlinear stochastic functional of a homogeneous Gaussian random process in a series of stochastic functionals (multiple Wiener integrals) that are orthogonal to each other in the ensemble mean sense. With various notations and definitions, many authors have applied the Wiener–Hermite expansion technique to several nonlinear stochastic problems in mathematical physics and engineering, such as the analysis and identification of nonlinear systems for noise input,[2,5,6–9] theory of turbulence,[10–12] wave propagation in a random medium,[13–15] and wave scattering from randomly rough surfaces.[16–20]

In a certain application[21,22] of the technique, however, it is necessary to represent a product of two random functions, each of which is described by Wiener–Hermite expansion, in terms of a Wiener–Hermite expansion. This can be done, in principle, by use of Imamura's symbolic calculus,[23,24] but no explicit formulas have been given in the literature. In this paper, we present a new mathematical formula for decomposing a product of two different multiple Wiener integrals into a series of Wiener–Hermite expansions. As a direct application of the decomposing formula, we then derive another formula developing a product of two random variables, given by Wiener–Hermite expansions, into a series of Wiener–Hermite expansions.

II. MULTIPLE WIENER INTEGRALS

According to Ref. 4, we briefly describe the definitions and notations concerning the Wiener–Hermite expansion, because various notations are available.[1–5,23]

Wiener–Hermite differentials: Let $B(x,\omega)$ be a real Brownian motion process on the real x axis $R = (-\infty,\infty)$, where ω is a probability parameter describing a sample point in the sample space Ω. However, we will often drop the probability parameter ω in equations below. The differential $dB(x,\omega) = B(x+dx,\omega) - B(x,\omega)$ is a strictly homogeneous Gaussian process with

$$\langle dB(x,\omega)\rangle = 0,$$
$$\langle dB(x_1,\omega)dB(x_2,\omega)\rangle = \delta(x_1-x_2)dx_1\,dx_2, \tag{1}$$

where the angular brackets denote the ensemble average over the sample space Ω, and $dB(x,\omega)/dx$ is the so-called Gaussian white noise. We assume, however, that the sample space Ω is of function space type,[25] where Ω is an infinite-dimensional Euclidian space and a sample function $dB(x,\omega)$ is projected to an infinite-dimensional vector ω in Ω in such a way that ω_x, the x component of ω, is given by $\omega_x = dB(x,\omega)$. Thus a function $g(\omega)$ of ω is always regarded as a functional of $dB(x,\omega)$, i.e., $g(\omega) = g[dB(\cdot,\omega)]$. Furthermore, a shift of a sample function by a, $dB(x,\omega)\to dB(x+a,\omega)$, induces a shift T^a of ω vector in Ω, namely, $dB(x+a,\omega) = dB(x,T^a\omega)$. In the case of the Brownian motion process,[2] the shift T^a is a measure-preserving transformation with a group property: $T^0 = 1$ (identity); $T^aT^b = T^{a+b}$. Once the measure-preserving transformation T is so defined, $g(T^x\omega) = g[dB(\cdot+x,\omega)]$ becomes a strictly homogeneous random function for any random variable $g(\omega) = g[dB(\cdot,\omega)]$.[26]

We introduce Wiener–Hermite differentials $h^{(m)}[dB(x_1),dB(x_2),\ldots,dB(x_m)]$, $m = 0,1,2,\ldots$, associated with the differential as

$$h^{(0)} = 1,\quad h^{(1)}[dB(x)] = dB(x),$$
$$h^{(2)}[dB(x_1),dB(x_2)] = dB(x_1)dB(x_2) - \delta(x_1-x_2)dx_1\,dx_2. \tag{2}$$

Higher-order Wiener–Hermite differentials may be obtained by the Rodrigues formula[4,23] or the recurrence formula[4]:

$$\begin{aligned}h^{(m)}&[dB(x_1),dB(x_2),\ldots,dB(x_m)]h^{(1)}[dB(x)]\\ &= h^{(m+1)}[dB(x_1),dB(x_2),\ldots,dB(x_m),dB(x)]\\ &\quad+\sum_{i=1}^{m} h^{(m-1)}[dB(x_1),dB(x_2),\ldots,dB(x_{i-1}),\\ &\quad dB(x_{i+1}),\ldots,dB(x_m)]\delta(x_i-x)dx_i\,dx\\ &\quad(m=1,2,3,\ldots),\end{aligned} \tag{3}$$

where m is an arbitrary positive integer. The Wiener–Hermite functionals enjoy the orthogonality relation[4,23,24]

$$\begin{aligned}\langle h^{(n)}&[dB(x_{i_1}),dB(x_{i_2}),\ldots,dB(x_{i_n})]\\ &\times h^{(m)}[dB(x_{j_1}),dB(x_{j_2}),\ldots,dB(x_{j_m})]\rangle\\ &= \delta_{mn}\delta_{ij}^{(m)}dx_{i_1}\,dx_{i_2}\cdots dx_{i_n}\,dx_{j_1}\,dx_{j_2}\cdots dx_{j_m}\\ &\quad(n,m=0,1,2,\ldots),\end{aligned} \tag{4}$$

where $\delta_{ij}^{(m)}$ stands for the sum of all distinct products of m delta functions of the form $\delta(x_{i_\alpha}-x_{j_\beta})$, $i=(i_1,i_2,\ldots,i_m)$, $j=(j_1,j_2,\ldots,j_m)$, all i_α and j_β appearing just once in each product, for example,

$$\delta_{ij}^{(2)} = \delta(x_{i_1}-x_{j_1})\delta(x_{i_2}-x_{j_2}) + \delta(x_{i_2}-x_{j_1})\delta(x_{i_1}-x_{j_2}). \tag{5}$$

Setting $n = 0$ in (4) and using (2), we find that $h^{(m)}$'s have zero averages except for $m = 0$,

$$\langle h^{(m)}[dB(x_1),dB(x_2),...,dB(x_m)]\rangle = \delta_{m0} \quad (m = 0,1,2,...). \quad (6)$$

Multiple Wiener integrals: For an m-variable function $F_m(x_1,x_2,...,x_m)$ belonging to $L^2(R^m)$, where $L^2(R^m)$ is the totality of square summable functions over the m-dimensional Euclidian space R^m, we define the m-tuple Wiener integral $I_m[F_m,\omega]$ as

$$I_m[F_m,\omega] = \int_{R^m} F_m(x_1,x_2,...,x_m) \times h^{(m)}[dB(x_1),dB(x_2),...,dB(x_m)] \quad (m = 0,1,2,...), \quad (7)$$

which is a random variable defined as a nonlinear stochastic functional of $dB(x,\omega)$. However, $I_0[F_0,\omega]$ is a deterministic constant equal to F_0. The integral kernel F_m will be referred to as the coefficient of the Wiener integral below. The Wiener integral (7) satisfies the identity holding in the ensemble mean square sense,

$$I_m[F_m,\omega] = I_m[F^s_m,\omega], \quad (8)$$

where F^s_m is the symmetric function defined by

$$F^s_m(x_1,x_2,...,x_m) = \frac{1}{m!}\sum_{(i)} F_m(x_{i_1},x_{i_2},...,x_{i_m}), \quad (9)$$

$(i) = (i_1,i_2,...,i_m)$ running all permutations of $(1,2,3,...,m)$. Because of (8) we often assume that the coefficient F_m is symmetrical. By (4), we get the orthogonality relation

$$\langle I_m[F_m,\omega]I_n[F_n,\omega]\rangle = m!\delta_{mn}\int_{R^m} F^2_m(x_1,x_2,...,x_m)dx_1\,dx_2\cdots dx_m, \quad (10)$$

for a symmetrical coefficient F_m.

Orthogonal development of a stochastic functional: We denote by $g(\omega)$ a stochastic functional of the differential $dB(x,\omega)$. If the functional $g(\omega)$ has a finite variance, i.e., $\langle|g(\omega)|^2\rangle < \infty$, it has the orthogonal development called the Wiener–Hermite expansion,

$$g(\omega) = \sum_{m=0}^{\infty} I_m[G_m,\omega], \quad (11)$$

which holds in the ensemble mean square sense. By (4) and (7) the coefficient G_m is an m-variable symmetric function given by the correlation

$$\langle g(\omega)h^{(m)}[dB(x_1),...,dB(x_m)]\rangle = m!G_m(x_1,...,x_m)dx_1\cdots dx_m. \quad (12)$$

The average and the mean square of $g(\omega)$ are easily calculated by the orthogonality relation (4) as

$$\langle g(\omega)\rangle = G_0, \quad (13)$$

$$\langle|g(\omega)|^2\rangle = \sum_{m=0}^{\infty} m!\int_{R^m}|G_m(x_1,...,x_m)|^2\,dx_1\cdots dx_m. \quad (14)$$

Thus the first term of the development (11) is equal to the average.

III. WIENER–HERMITE EXPANSION OF A PRODUCT OF TWO WIENER INTEGRALS

We present a new formula for Wiener–Hermite expansion of a product of two Wiener integrals. However, we first introduce the k-dimensional inner product to simplify the equations. For symmetric functions G_m and F_m, belonging to $L^2(R^m)$ and $m = 0,1,2,...$, we define the k-dimensional inner product by the relation

$$\{G_m,F_n\}_k = \int_{R^k} G_m(x_1,x_2,...,x_{m-k},y_1,y_2,...,y_k) \times F_n(X_1,X_2,...,X_{n-k},y_1,y_2,...,y_k)dy_1\,dy_2\cdots dy_k \quad [k\leqslant\min(n,m)], \quad (15)$$

which is a function of $n + m - 2k$ variables, but asymmetrical in general.

Theorem 1: A product of two multiple Wiener integrals with symmetrical coefficients can be represented by the sum of multiple Wiener integrals as follows:

$$I_m[G_m,\omega]\times I_n[F_n,\omega] = \sum_{k=0}^{\min(n,m)} \frac{m!n!}{(m-k)!k!(n-k)!} I_{m+n-2k} \times[\{G_m,F_n\}_k,\omega], \quad (16)$$

which is the main result in this paper.

Proof: We prove this by mathematical induction. Since (16) is symmetrical with m and n, we can assume $n\leqslant m$ and m is an arbitrary non-negative integer, without loss of generality. Because (16) is trivial for $n = 0$, let us prove (16) for $n = 1$ in the first step. Multiplying $G_m(x_1,x_2,...,x_m)F_1(x)$ to both sides of (3), and integrating them with respect to $x_1,x_2,...,x_m$ and x over R^{m+1}, we easily find

$$I_m[G_m,\omega]\times I_1[F_1,\omega] = I_{m+1}[\{G_m,F_1\}_0,\omega] + mI_{m-1}[\{G_m,F_1\}_1,\omega], \quad (17)$$

where G_m has been assumed to be symmetrical. Equation (17) is identical to (16) for $n = 1$. Next we assume that (16) is valid for all n less than m, where m is an arbitrary integer. Under this assumption we prove that (16) holds again for $n = n + 1$. Since $F_{n+1}(x_1,x_2,...,x_n,x_{n+1})$ is a symmetrical function belonging to $L^2(R^{n+1})$, then it is symmetrical with respect to the first n variables and is an element of $L^2(R^n)$ when the $(n+1)$th variable x_{n+1} is fixed. Therefore we can replace $F_n(x_1,x_2,...,x_n)$ by $F_{n+1}(x_1,x_2,...,x_n,x_{n+1})$ in (16), keeping x_{n+1} constant. Then we obtain

$$I_m[G_m,\omega]\int_{R^n}F_{n+1}(x_1,x_2,\ldots,x_n,x_{n+1})h^{(n)}[dB(x_1),\ldots,dB(x_n)]$$
$$=\sum_{k=0}^{n}\frac{m!n!}{(m-k)!k!(n-k)!}\int_{R^{n+m-2k}}\int_{R^k}G_m(x_1,x_2,\ldots,x_{m-k},y_1,y_2,\ldots,y_k)F_{n+1}(X_1,X_2,\cdot,X_{n-k},y_1,y_2,\cdot,y_k,x_{n+1})$$
$$\times dy_1\,dy_2\cdots dy_k\,h^{(n+m-2k)}[dB(x_1),\ldots,dB(x_{m-k}),dB(X_1),\ldots,dB(X_{n-k})]\,, \qquad (18)$$

which holds for any value of x_{n+1}. Multiplying $dB(x_{n+1})$ with both sides, integrating with x_{n+1} over R^1, and applying the recurrence formula (3) to the both sides again, we obtain

$$I_m[G_m,\omega]\times I_{n+1}[F_{n+1},\omega]+I_m[G_m,\omega]\times n\times I_{n-1}[Q_{n-1},\omega]$$
$$=\sum_{k=0}^{n}\frac{m!n!}{(m-k)!k!(n-k)!}I_{n+m+1-2k}[\{G_m,F_{n+1}\}_k,\omega]$$
$$+\sum_{k=0}^{n}\frac{m!n!(m-k)}{(m-k)!k!(n-k)!}I_{n+m-1-2k}[\{G_m,F_{n+1}\}_{k+1},\omega]$$
$$+\sum_{k=0}^{n}\frac{m!n!(n-k)}{(m-k)!k!(n-k)!}I_{n+m-1-2k}[\{G_m,Q_{n-1}\}_k,\omega], \qquad (19)$$

where the symmetry of F_{n+1} has been used again and we have set

$$Q_{n-1}(x_1,x_2,\ldots,x_{n-1})$$
$$=\int_{R^1}F_{n+1}(x_1,x_2,\ldots,x_{n-1},x_{n+1},x_{n+1})dx_{n+1}. \qquad (20)$$

Applying (16) for $n=n-1$ to calculate $I_m[G_m,\omega]\times I_{n-1}[Q_{n-1},\omega]$, we easily find that $n\times I_m[G_m,\omega]\times I_{n-1}[Q_{n-1},\omega]$ equals the third sum in the right-hand side in (19), because of the factor $(n-k)$ in (19). Thus we obtain the identity

$$I_m[G_m,\omega]\times I_{n+1}[F_{n+1},\omega]$$
$$=\sum_{k=0}^{n}\frac{m!n!}{(m-k)!k!(n-k)!}$$
$$\times I_{n+m+1-2k}[\{G_m,F_{n+1}\}_k,\omega]$$
$$+\sum_{k=0}^{n}\frac{m!n!(m-k)}{(m-k)!k!(n-k)!}$$
$$\times I_{n+m-1-2k}[\{G_m,F_{n+1}\}_{k+1},\omega]. \qquad (21)$$

Rewriting $k+1$ as k in the second sum, and using the identity

$$\frac{m!n!}{(m-k)!k!(n-k)!}+\frac{m!n!}{(m-k)!(k-1)!(n+1-k)!}$$
$$=\frac{m!(n+1)!}{(m-k)!k!(n+1-k)!}, \qquad (22)$$

we can easily add up the two summations in (21). As a result, we may find that (21) is identical with (16) for $n=n+1$.

As a direct application of Theorem 1, let us prove the following theorem.

Theorem 2: Let $g(\omega)$ and $f(\omega)$ be random variables described by Wiener–Hermite expansions:

$$g(\omega)=\sum_{m=0}^{\infty}I_m[G_m,\omega]\,, \qquad (23)$$

$$f(\omega)=\sum_{n=0}^{\infty}I_n[F_n,\omega]\,, \qquad (24)$$

where the coefficients G_m and F_n are symmetrical functions with their arguments. Then the product of $g(\omega)$ and $f(\omega)$ is given by another Wiener–Hermite expansion as

$$g(\omega)\times f(\omega)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\sum_{k=0}^{m}\frac{(n+m-k)!(n+k)!}{(m-k)!k!n!}$$
$$\times I_m[\{G_{n+m-k},F_{n+k}\}_n,\omega]\,. \qquad (25)$$

Proof: From (23), (24), and (16) we obtain

$$g(\omega)\times f(\omega)=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}I_m[G_m,\omega]\times I_n[F_n,\omega]$$
$$=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\sum_{k=0}^{\min(n,m)}\frac{n!m!}{(n-k)!k!(m-k)!}$$
$$\times I_{n+m-2k}[\{G_m,F_n\}_k,\omega]\,, \qquad (26)$$

where the set of the three-dimensional lattice points (n,m,k) in the triple sum will be called the T set, which is illustrated in Fig. 1. Let us arrange the triple sum in the right-hand side to get (25). This can be done easily with help of Fig. 1. In the first step, we set

$$n+m=l+2k, \qquad (27)$$

where l runs over all non-negative integers because $0\leqslant k\leqslant\min(m,n)$. Next we consider two-dimensional vectors (m,n) satisfying (27). For any given k and l, the two-dimensional vector (m,n) takes only $(l+1)$ different values,

$$(m,n)=(l+k,k),(l+k-1,k+1),\ldots,(k,l+k),$$
$$=\{(l+k-p,k+p);p=0,1,2,\ldots,l\}, \qquad (28)$$

which satisfy $0\leqslant k\leqslant\min(m,n)$ and belong to the T set. [However, for example, $(m,n)=(0,2k+l)$ does not satis-

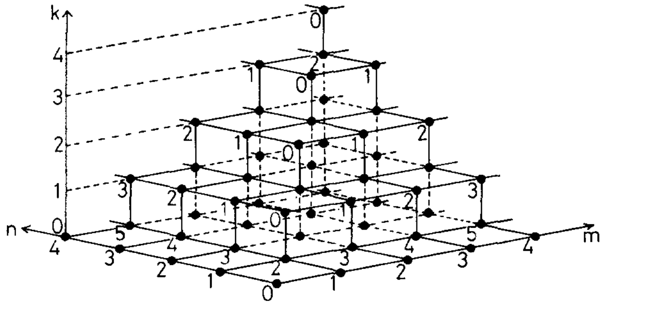

FIG. 1. Three-dimensional lattice points (m,n,k) in the T set. Dots with 0 satisfy $m+n-2k=0$, dots with 1 enjoy $n+m-2k=1$, and so on.

fy $0 \leqslant k \leqslant \min(m,n)$ in general, and hence it is not an element of the T set.] Therefore, setting $m = l + k - p$ and $n = k + p$, we may rearrange the triple sum in (26) as

$$g(\omega)\times f(\omega) = \sum_{l=0}^{\infty}\sum_{k=0}^{\infty}\sum_{p=0}^{l}\frac{(k+l-p)!(k+p)!}{(l-p)!p!k!} \times I_l[\{G_{k+l-p},F_{k+p}\}_k,\omega], \quad (29)$$

which is the same equation as (25).

IV. EXAMPLES

Example 1: When $g(\omega)$ and $f(\omega)$ are given by (23) and (24), respectively, let us calculate the average value of $g(\omega)\times f(\omega)$ by the use of (25). When $m\neq 0$, the m-tuple Wiener integrals in (25) have zero averages,

$$\langle I_m[\{G_{n+m-k},F_{n+k}\}_n,\omega]\rangle = 0 \quad (m\neq 0), \quad (30)$$

because the Wiener–Hermite differentials have zero averages by (6). Therefore, averaging both sides of (25), and interchanging the order of summation and averaging, we obtain

$$\langle g(\omega)\times f(\omega)\rangle = \sum_{n=0}^{\infty} n! I_0[\{G_n,F_n\}_n,\omega] = \sum_{n=0}^{\infty} n!\{G_n,F_n\}_n, \quad (31)$$

which can be obtained directly from (23), (24), and the orthogonality relation (4).

Example 2: Let $f(\omega)$ and $g(\omega)$ be random variables given by exponentials of single Wiener integrals

$$g(\omega) = \exp\left[\int_R G(x)dB(x)\right] = \sum_{n=0}^{\infty} I_m[\underline{G}_m,\omega], \quad (32)$$

$$f(\omega) = \exp\left[\int_R F(x)dB(x)\right] = \sum_{n=0}^{\infty} I_n[\underline{F}_n,\omega], \quad (33)$$

where coefficients $\underline{G}_m$ and $\underline{F}_m$ are given by[2,4]

$$\underline{G}_m(x_1,x_2,\ldots,x_m) = \frac{1}{m!}\exp\left[\frac{1}{2}\int_R G^2(x)dx\right]G_m(x_1,x_2,\ldots,x_m),$$

$$G_m(x_1,x_2,\ldots,x_m) = G(x_1)G(x_2)\cdots G(x_m), \quad (34)$$

$$\underline{F}_n(x_1,x_2,\ldots,x_n) = \frac{1}{n!}\exp\left[\frac{1}{2}\int_{R^1} F^2(x)dx\right]F_n(x_1,x_2,\ldots,x_n),$$

$$F_n(x_1,x_2,\ldots,x_n) = F(x_1)F(x_2)\cdots F(x_n). \quad (35)$$

Replacing $G(x)$ in (32) with $[G(x)+F(x)]$, we obtain the Wiener–Hermite expansion of $g(\omega)\times f(\omega)$ as

$$g(\omega)f(\omega) = \exp\left\{\int_R [G(x)+F(x)]dB(x)\right\} = \sum_{m=0}^{\infty} I_m[(\underline{G+F})_m,\omega], \quad (36)$$

where

$$(\underline{G+F})_m = \frac{1}{m!}\exp\left\{\frac{1}{2}\int_R [G(x)+F(x)]^2dx\right\}(G+F)_m, \quad (37)$$

$$(G+F)_m = [G(x_1)+F(x_1)][G(x_2)+F(x_2)]\times\cdots\times[G(x_m)+F(x_m)]. \quad (38)$$

Now let us derive (36) from Wiener–Hermite expansions for $g(\omega)$ and $f(\omega)$ as an application of Theorem 2. If we apply (25) to the product of $f(\omega)$ and $g(\omega)$, we obtain

$$f(\omega)\times g(\omega) = \left(\sum_{m=0}^{\infty} I_m[\underline{F}_m,\omega]\right)\left(\sum_{m=0}^{\infty} I_m[\underline{G}_m,\omega]\right) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\sum_{k=0}^{m}\frac{(n+m-k)!(n+k)!}{(m-k)!k!n!} \times I_m[\{\underline{G}_{n+m-k},\underline{F}_{n+k}\}_n,\omega]. \quad (39)$$

Now let us calculate the right-hand side of (39). By (34) and (35), one easily finds

$$\sum_{n=0}^{\infty}\sum_{k=0}^{m}\frac{(n+m-k)!(n+k)!}{(m-k)!k!n!}I_m[\{\underline{G}_{n+m-k},\underline{F}_{n+k}\}_n,\omega] = \sum_{n=0}^{\infty}\frac{1}{n!}\exp\left\{\frac{1}{2}\int[G^2(x)+F^2(x)]dx\right\}[\{G,F\}_1]^n\sum_{k=0}^{m}\frac{1}{(m-k)!k!}I_m[\{G_{m-k},F_k\}_0,\omega]$$
$$= \exp\left\{\frac{1}{2}\int[G(x)+F(x)]^2dx\right\}\sum_{k=0}^{m}\frac{1}{(m-k)!k!}I_m[\{G_{m-k},F_k\}_0,\omega]. \quad (40)$$

On the other hand, decomposing the product in the right-hand side of (38) and using (7), we obtain

$$I_m[(G+F)_m,\omega] = \sum_{k=0}^{m}\frac{m!}{(m-k)!k!}I_m[\{G_{m-k},F_k\}_0,\omega]. \quad (41)$$

Equations (40) and (41) given an identity

$$\sum_{n=0}^{\infty}\sum_{k=0}^{m}\frac{(n+m-k)!(n+k)!}{(m-k)!k!n!}I_m[\{\underline{G}_{n+m-k},\underline{F}_{n+k}\}_n,\omega] = \frac{1}{m!}\exp\left\{\frac{1}{2}\int_R[G(x)+F(x)]^2dx\right\} \times I_m[(G+F)_m,\omega] = I_m[(\underline{G+F})_m,\omega], \quad (42)$$

where (37) was used to get the second equality. This relation means that the right-hand side of (39) is identical to (36).

V. CONCLUSIONS

We have presented a new formula for decomposing a product of two multiple Wiener integrals into series of Wiener–Hermite expansions. The formula was proved in a formal way, but no rigorous mathematical proof was given.

Simple examples were described as an application of the formula. Further applications of the formulas, however, will be published elsewhere.

[1]R. H. Cameron and W. T. Martin, "The orthogonal development of nonlinear functionals in series of Fourier–Hermite functionals," Ann. Math. **48**, 385 (1947).

[2]N. Wiener, *Nonlinear Problems in Random Theory* (MIT, Cambridge, MA, 1958).

[3]K. Itô, "Multiple Wiener integrals," J. Math. Soc. Jpn. **3**, 157 (1851); "Complex multiple Wiener integrals," Jpn. J. Math. **22**, 63 (1952).

[4]H. Ogura, *Stochastic Processes* (Corona, Tokyo, 1978) (in Japanese).

[5]W. J. Rugh, *Nonlinear System Theory* (John Hopkins U. P., Baltimore, MD, 1981).

[6]J. F. Barrett, "Hermite functional expansion and the calculation of output auto-correlation and spectrum for any time invariant nonlinear system with noise input," J. Electron. Control **16**, 107 (1963).

[7]S. Yasui, "Stochastic functional Fourier series, Volterra series and nonlinear system analysis," IEEE Trans. Autom. Control **24**, 230 (1979).

[8]M. Schtzen, "Nonlinear system modeling based on the Wiener theory," Proc. IEEE **69**, 1557 (1981).

[9]E. Isobe and S. Sato, "On the identification of a class of nonlinear systems using the Wiener's method," IECE Jpn. Trans. **J64-A**, 1026 (1981).

[10]A. Siegel, T. Imamura, and W. C. Meecham, "Wiener–Hermite expansion in model turbulence in the late decay stage," J. Math. Phys. **6**, 707 (1965).

[11]H. D. Gogge and W. C. Meecham, "The Wiener–Hermite expansion applied to decaying isotropic turbulence using a renormalized time-dependent base," J. Fluid Mech. **85**, 325 (1978).

[12]R. Kambe, M. Doi, and T. Imamura, "Turbulent flows near flat plates," J. Phys. Soc. Jpn. **49**, 763 (1980).

[13]H. Ogura, "Theory of waves in a homogeneous random medium," Phys. Rev. A **11**, 942 (1975).

[14]H. Ogura and J. Nakayama, "Initial-value problem of the one-dimensional wave propagation in a homogeneous random medium," Phys. Rev. A **11**, 957 (1975).

[15]H. Ogura and Y. Yoshida, "Wave propagation in a slightly lossy random medium," Phys. Rev. A **14**, 796 (1976).

[16]J. Nakayama, H. Ogura, and M. Sakata, "Scattering of a scalar wave from a slightly random surface," J. Math. Phys. **22**, 471 (1981).

[17]J. Nakayama, H. Ogura, and M. Sakata, "A probabilistic theory of electromagnetic wave scattering from a slightly random surface: 1, Horizontal polarization," Radio Sci. **16**, 831 (1981).

[18]J. Nakayama, "Anomalous scattering from a slightly random surface," Radio Sci. **17**, 558 (1982).

[19]J. Nakayama, K. Mitzutani, and M. Tsuneoka, "Scattering of electromagnetic waves from a perfectly conductive slightly random surface: Depolarization in backscatter," J. Math. Phys. **27**, 1436 (1986).

[20]J. Nakayama, "Scattering from a random rough surface: Linear equations for coefficients of Wiener–Hermite expansion of the wave field," Radio Sci. **21**, 707 (1986).

[21]M. Tsuneoka, M.E. thesis, Kyoto Institute of Technology, 1987.

[22]E. Ohmori, B.E. thesis, Department of Electronics, Kyoto Institute of Technology, 1987.

[23]T. Imamura, W. C. Meecham, and A. Siegel, "Symbolic calculus of the Wiener process and Wiener–Hermite functionals," J. Math. Phys. **6**, 695 (1965).

[24]T. Imamura, *Mathematics in Random Field Theory* (Iwanami-Shoten, Tokyo, 1976) (in Japanese).

[25]J. L. Doob, *Stochastic Processes* (Wiley, New York, 1953).

[26]The Wiener–Hermite expansion is often used to represent a homogeneous random function. This paper, however, deals with only random variables represented by Wiener–Hermite expansions. This is because a homogeneous random function can be obtained from a random variable by the measure-preserving transformation.

Spontaneous splitting and internal isometries of superstring vacua

Brett McInnes
Department of Mathematics, National University of Singapore, 10 Kent Ridge Crescent 0511, Republic of Singapore

(Received 29 September 1986; accepted for publication 15 July 1987)

Superstring vacua are normally presumed to be of the form $M \times K$, where $\dim(M) = 4$, $\dim(K) = 6$, and where $\times$ denotes the *global* Riemannian product. Since, however, one would ultimately wish to understand the external/internal distinction in terms of some dynamical mechanism ("spontaneous splitting") involving vacuum expectation values of local fields, it may be preferable to use a *local* Riemannian product at the outset. Here it is shown that these spaces, which have the same local (block-diagonal) type of metric as $M \times K$, can be described and classified by examining the isometry groups of the Calabi–Yau manifolds which have been proposed as models for the internal superstring vacua.

I. SUPERSTRINGS AND CONTINUOUS ISOMETRIES

Much recent work on superstring phenomenology[1] is based on vacuum models of the form $M \times K$, where M is four dimensional and flat, while K is six dimensional and Ricci flat. This represents a radical departure from previous Kaluza–Klein formulations, in which the internal space was always chosen to have a large group of continuous isometries. For, as is well known, the condition of Ricci flatness strongly restricts the isometry group of a compact Riemannian manifold. In fact, the internal manifolds considered in the superstring context [complex manifolds with SU(3) holonomy group] have no continuous isometries whatever. The only continuous isometries of the vacuum are those possessed by the external space. It is as if—speaking figuratively—the external space is formed from the Killing vectors which have been forced to "split off" from the internal space.

This rather strange state of affairs—a highly symmetric external space coupled with a highly asymmetric internal space—does not, in the usual approach, appear to spring from any deeper principle; it merely arises as a natural consequence of the analysis. However, it seems unlikely that this striking feature of the model can be of no consequence. We have argued elsewhere[2] that, on the contrary, this observation may provide the key to an understanding of the origin of the internal/external dichotomy itself. The remainder of this section consists of a brief explanation of this remark.

For the sake of convenience (and because path-integral quantization will eventually be necessary, though that is beyond our present scope) we shall use a Euclidean background. Now general relativity would lead us to expect that a suitable vacuum for a ten-dimensional theory should be a flat ten-dimensional manifold with as much symmetry as the topology permits—indeed, one might go further and suggest that this topology should be that of $\mathbb{R}^{10}$. Evidently the actual vacuum is less symmetrical than this, in several senses. Experience has shown that this is not necessarily an unsatisfactory feature of a theory, *provided* that the symmetries are broken "spontaneously" rather than "by hand." By this, one means that the various asymmetries of the vacuum should arise, through some mechanism, as a consequence of the nature and behavior of the vacuum expectation values of physical fields. Otherwise, the breakdown of symmetry would introduce an intolerable degree of arbitrariness into the theory.

One of the more conspicuous asymmetries of the superstring vacuum is the "splitting" of the manifold into internal and external factors. Can this asymmetry be explained in a spontaneous manner? The difficulty here is that the product is normally interpreted, either implicitly or explicitly, as a *global* product: technically, one postulates the existence of a global isometric mapping from the ten-dimensional space P to $M \times K$. Since any spontaneous mechanism involves expectation values of *locally* defined fields, it is difficult to see how such a mechanism can generate a global splitting. Later we shall see that, in the context of very particular superstring models, this problem can be solved; but in a more general context, the above remarks suggest that (at the outset) we should aim to induce only a *local* splitting. That is, M and K should be obtained as submanifolds of P which intersect orthogonally at every point (so that the metric has the usual block-diagonal form in every adapted coordinate patch) but which cannot necessarily be extended to yield a global product. The problem of determining the additional global conditions which must be imposed in order to obtain a global product can then be studied separately.

We return, then, to the full ten-dimensional space P, and attempt to induce a local splitting in a spontaneous manner. We assume that P is Ricci flat and compact. (The compactness should also arise spontaneously; but that is a familiar problem, for which well-known techniques exist.[3]) Thus we are supposing that compactification precedes splitting. Next, assume that P admits a set of linearly independent, nonzero, covariant-constant vector fields. In general, Riemannian manifolds do not readily admit constant vector fields. But on Ricci-flat compact Riemannian manifolds, such fields exist in abundance: according to a series of theorems due to Yano–Bochner,[4] and others, whole classes of vector fields are necessarily constant on such manifolds. For example, this is true of all Killing fields, all harmonic vector fields, and so on. It can also be shown[2] that any vector field on such a manifold which satisfies Maxwell's equations (plus a gauge fixing condition) is constant, and of course the

 0022-2488/87/112564-05$02.50

same will be true of gauge fields corresponding to a maximal torus of any gauge group. Hence, it may not be too much to hope that these fields can be generated spontaneously, in the sense defined earlier.

Now according to the local version of the de Rham decomposition theorem (see Ref. 5, p. 185; for a more detailed explanation, with examples, see Ref. 2), the existence of constant vector fields on P will induce the latter to split into a *local* Riemannian product of a *flat* manifold M with a Ricci-flat manifold K. Since every Killing field is covariant constant, M will be maximally symmetric (to the extent permitted by the topology) while K will possess no continuous isometries whatever. Thus our formulation, oversimplified though it undoubtedly is, can explain not only the internal/external split, but also the total absence of continuous isometries from the internal space. That is, we propose to identify M, the flat space generated by the Killing vectors, with external space, and to regard K as internal space. (Note that every constant vector field is a Killing field, and so we can refer to our fields as such even if they originate in some quite different way.)

In short, at the *local* level, we have obtained precisely the standard vacuum structure: the metric of P has the usual block-diagonal form, with one block pertaining to a flat space, the other to a Ricci-flat space devoid of continuous isometries. But now this local splitting can arise spontaneously, as it must if we are ever to give a convincing account of our failure to observe the internal space directly. However, the global structure can differ from the standard picture, as we shall now explain.

II. THE GLOBAL STRUCTURE

The global structure that arises from the spontaneous splitting can differ from the usual framework in two distinct ways.

First, by means of the de Rham theorem and the Hopf–Rinow theorem,[6] it can be shown[2] that *both* the internal and the external spaces are necessarily compact. [This is not obvious, since we do not (yet) know whether P is a topological product of M with K.] In Ref. 1, it is shown that M must be flat, but the question of its topology is left open: it could be either compact or noncompact. In the present case, this option no longer exists. (If there are objections to the compactification of time, then one can interpret M as three-dimensional space; unfortunately, we are unable to specify the dimensionality of M, since that will be determined by the details of the particular mechanism used to generate the constant vector fields.)

Second, and far more seriously, we cannot always ensure that the local splitting obtained above will extend to a global splitting. The global version of the de Rham decomposition theorem (Ref. 5, p. 187) states that the extension can be performed if the manifold is complete and simply connected. Unfortunately, however, while P is complete (since it is compact), it is certainly not simply connected. For if it were, then it would split globally into a product of simply connected factors, so that M would be flat and simply connected. This is impossible, since M is compact. Thus the problem of extending the decomposition cannot be approached in this way.

If the extension cannot be performed, then P has a structure roughly analogous to that of the *flat* Möbius strip (that is, the space obtained from the unit square in $\mathbb{R}^2$ by means of certain topological identifications, *not* the more familiar Möbius surface in $\mathbb{R}^3$). This space is a local metric product; yet it contains pairs of submanifolds that are everywhere orthogonal but that intersect twice. In the case of P, this would mean (see below) that each internal space could intersect external space several times. The distance between two given points in the external space can differ enormously, depending on whether the route is taken through the internal space.[2]

Despite its somewhat bizarre aspect, it is not very clear whether this effect would be readily detected. At the level of principle, however, this kind of "wormholelike" behavior could lead to difficulties with causality, since it is evidently possible to transmit signals "faster than light" through internal space. Of course, the same objection could be—indeed, has been[7]—raised against true wormholes. In each case, it is possible to reply that the special-relativistic formulation of causality should not be foisted on spaces with structures more complicated than that of Minkowski space. But in view of the fact that we are dealing with a *vacuum* model, these additional complications may not be welcome.

We shall not attempt to settle this question here; instead, we return to the problem of determining the conditions under which the local splitting of the metric *can* be extended to the global level.

III. GLOBAL SPONTANEOUS SPLITTING

As was pointed out earlier, the covariant-constant vector fields which generate M can be interpreted as Killing fields of P. Thus they can be regarded as arising from the action of a (necessarily compact) group of isometries G. It may be shown[2] that G is Abelian and that P has the structure of a principal fiber bundle (P, Q, G), where the base manifold Q is (by definition) P/G, and in fact $G = \mathrm{U}(1)^m$ (where $m = \dim M$, that is, 3 or 4, depending on whether time is included). We stress that this bundle is *in no way* related to the familiar gauge bundles.[8] Here it is the *external* space which is to be identified with the fibers; the internal spaces appear as holonomy bundles (see below), whereas Q has (in general) no physical significance whatever. Note that the topology of M is now fixed: it is that of a torus.

Now by a well-known construction,[9,10] the existence of a G-invariant metric on the bundle space of a principal bundle defines a connection: the horizontal distribution can be defined to be orthogonal to the vertical distribution. Applying this in our case, we obtain a natural connection Γ on (P, Q, G). As usual, the curvature of Γ measures the failure of the horizontal distribution to be involutive. But here the horizontal distribution *is* involutive; indeed, by the local de Rham theorem, it integrates to yield the internal spaces. Thus Γ is flat, and it can be shown[2] that the internal spaces are just the holonomy bundles[5] of Γ.

By the holonomy reduction theorem,[5] the internal spaces are principal bundles over Q with structural group H, the holonomy group of Γ. As is well known (see Ref. 1 and below) flat connections on principal bundles need not have trivial holonomy groups H: rather, H may be a discrete subgroup of the structural group. In our case, $G = \mathrm{U}(1)^m$, so H can be any product of m cyclic groups, $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_m}$. Of course, H *can* be trivial; this will occur, for example, if Q happens to be simply connected. In that case each holonomy bundle will be isomorphic to Q, and P becomes a global product of M and Q (in the Riemannian, as well as the topological sense).

But if H is not trivial, then each internal space must be interpreted as a bundle with a discrete structural group. A given holonomy bundle will intersect each vertical space of P in as many points as there are elements in H (that is, parallel transport around closed loops in Q has effects controlled by H). Since H, being a subgroup of G, acts isometrically, these points of intersection will be evenly distributed. In short, we have precisely the situation described in general terms in the preceding section. Each internal space (holonomy bundle) intersects every external (vertical) space orthogonally, and yet each given pair intersects at more than one point. Thus we arrive at an important conclusion: it is the holonomy group H that obstructs the extension of the local (spontaneous) splitting to the global level.

We can formulate the above remark in a more useful way as follows. Since each holonomy bundle is contained in P, and since H is contained in G, it follows that H acts isometrically on the holonomy bundles. Thus H is a *subgroup* of the isometry group of the internal space.

Now consider all this in the particular context of the superstring vacuum. According to our construction, the internal space K is devoid of *continuous symmetries*. But—and this is now a crucial point—K may still retain a *discrete* isometry group. Since the isometry group of a compact manifold is itself compact, this discrete group must in fact be finite. We now obtain a very simple formulation of the problem: the obstruction to extending the local de Rham decomposition to a global splitting resides in the cyclic subgroups of the finite isometry group of K. In other words, the problem of the "multiple intersections" arises from the residual discrete symmetries which remain with the internal space after external space "splits off."

All this casts an interesting light on the models considered in Ref. 1. For example, consider the space $Y_{4,5}$ defined by the equation

$$z_1^5 + z_2^5 + z_3^5 + z_4^5 + z_5^5 = 0$$

in $\mathbb{C}P^4$. This manifold admits a Ricci-flat metric, with respect to which it has no continuous isometries. The manifold is mapped into itself by a certain $\mathbb{Z}_5 \times \mathbb{Z}_5$ group of transformations. Although the metric is not known explicitly, it follows from Calabi's uniqueness theorem[11] that this group acts isometrically. Evidently there is some danger that, in this case, multiple intersections will arise either from $\mathbb{Z}_5 \times \mathbb{Z}_5$ itself or from one of its subgroups. However, it is far from clear that this danger can actually materialize: while we have shown that the existence of a cyclic subgroup of the isometry group of K is *necessary* for the existence of multiple intersections, we have not shown that it is *sufficient* to permit the existence of such spaces. In order to clarify the situation, we must answer the following question: given a compact Riemannian manifold K and a cyclic subgroup of its isometry group, is it always possible to construct a vacuum model in which K has multiple intersections with external space? This question is answered by the following result.

Proposition: Let K be a connected compact Riemannian manifold with $\dim(K) > 1$, let G be a compact Abelian Lie group, and let H be any finite subgroup of G that acts isometrically and freely on K. Then there exists a principal G-bundle P over K/H and a G-invariant metric on P such that P admits a connection Γ with holonomy bundles isometric to K and holonomy group isomorphic to H.

Proof (Sketch): As a subgroup of G, H has a natural left action on G. Thus if we regard K as a principal H-bundle over K/H, then we can construct the associated bundle $P = [K \times G]/H$, with standard fiber G and base manifold K/H. But now P in its turn admits a right action by G, given by $\bar{g}$: $\{k,g\} \to \{k,g\bar{g}\}$, where $\{k,g\}$ denotes an equivalence class in $[K \times G]/H$. Since H acts freely on K, it follows that G acts freely on P, and in fact P now becomes a principal G-bundle over K/H. The mapping $k \to \{k,e\}$, where $k \in K$ and e is the identity element of G, allows us to display $(K, K/H, H)$ as a reduced subbundle of $(P, K/H, G)$.

A deep theorem of Nomizu[12] states that for any principal bundle $(K, K/H, H)$, where K is connected, K/H is paracompact, and $\dim(K/H) > 1$, there exists a connection with holonomy bundles that coincide with K. In our case K and K/H satisfy all of these conditions, since H is finite. Let Γ be the corresponding connection; evidently its holonomy group is H. Now this connection extends, according to the connection mapping theorem,[15] to a connection $\bar{\Gamma}$ on $(P, K/H, G)$. The holonomy group of $\bar{\Gamma}$ is still H, and its holonomy bundles are isomorphic to K. (Both Γ and $\bar{\Gamma}$ are flat, since H is discrete, that is, its Lie algebra is trivial.)

The manifold K/H inherits a natural metric from K. By means of a standard construction,[9,10] one can use this metric, together with $\bar{\Gamma}$ and any invariant metric on G, to define a G-invariant metric on P. It is clear that, with respect to this metric, the holonomy bundles of $\bar{\Gamma}$ are isometric to K. This completes the proof.

In physical language, this result means that if the isometry group of a Ricci-flat manifold K contains a cyclic subgroup that acts freely, then it will certainly be possible to construct vacuum models in which K intersects external space more than once. (In principle, the proof yields an explicit construction; unfortunately, however, the proof of Nomizu's theorem is highly technical, and so we shall not discuss the procedure here.) The various possibilities are classified by the distinct cyclic subgroups of the isometry group. For example, in the case of $Y_{4,5}$ the isometry group itself is $\mathbb{Z}_5 \times \mathbb{Z}_5$, which can be regarded as a subgroup of $\mathrm{U}(1)^m$. As 5 is prime, the distinct subgroups that act freely are $\mathbb{Z}_5 \times \mathbb{Z}_5$, the two different $\mathbb{Z}_5$ groups, and 1, the group consisting of one element. In this last case, and only in this case, we obtain a true Riemannian product. But this is only one of several possibilities. At the other end of the spectrum, it is possible to construct a model with the *same* local metric but in which each internal space intersects external space 25 times. Unless we simply eliminate these other possibilities by fiat, which would subvert the whole philosophy of our approach, we have no grounds *a priori* to expect that the usual model will be produced.

Since our problems arise from the $\mathbb{Z}_n$ subgroups of the isometry group of K, it is natural to search for examples of K with isometry groups having no such subgroups. But every finite group, with the exception of 1, has a nontrivial $\mathbb{Z}_n$ subgroup. For if g is some element of the group not equal to the identity, then the subgroup generated by g will be isomorphic to $\mathbb{Z}_n$ for some n. Combining this with the proposition proved earlier, we see that the *only* way to eliminate multiple intersections right from the outset is to choose K to have no freely acting groups of isometries, either continuous or discrete. Now of course if F is the maximal freely acting finite group of isometries of K, then we can directly eliminate F by considering K/F, which will also be a manifold. But this again is no more than an *ad hoc* device, *unless* we can find independent physical reasons for preferring K/F to K. Remarkably, however, such reasons do exist and have been discussed extensively in the literature.[1,13]

Consider, for example, $K = Y_{4,5}$. Here we find $F = \mathbb{Z}_5 \times \mathbb{Z}_5$, and we wish to ask whether there are physical reasons for preferring $Y = Y_{4,5}/(\mathbb{Z}_5 \times \mathbb{Z}_5)$ to $Y_{4,5}$ itself. The answer is well known, and is given in Ref. 1. First, the Euler characteristic of Y is much smaller than that of $Y_{4,5}$, a fact that permits models based on Y— but not $Y_{4,5}$— to make realistic predictions regarding the number of generations of particles. Second, Y, unlike $Y_{4,5}$, is multiply connected; this permits the implementation of a novel and highly attractive method of gauge symmetry breaking (based, once again, on the fact that the holonomy group of a flat connection need not be trivial if the base manifold is multiply connected). More generally, it appears to be the rule that Calabi–Yau manifolds with few or no freely acting symmetries yield realistic models more readily than those with many.[13,14]

To summarize, then, we have the following situation. Spontaneous splitting, being essentially a local mechanism, can only yield a local splitting of the manifold and its metric. The question of whether the splitting can be extended globally depends on the freely acting $\mathbb{Z}_n$ subgroups of the isometry group of the internal space. But in many instances—in practice, one would need to investigate this on a case-by-case basis—there are good physical reasons to factor out these subgroups. In such a case, the splitting can be extended, and the equation $P = M \times K$ is strictly valid in both the Riemannian and the topological senses. In short, we have obtained a *global* spontaneous splitting.

We conclude this section with a technical remark. First, note that all of our discussion above relates to isometry groups that act freely. Now since holonomy groups are subgroups of the structure group of a principal bundle, they always act freely on holonomy bundles. Thus if a group H does not act freely on K, then it is certainly *not* possible to represent K as a holonomy bundle with H as holonomy group. Such isometry groups are therefore innocuous: they cannot give rise to multiple intersections, and our considerations provide no motive for factoring them out. (Note, however, that a group which does not act freely can contain a subgroup which does.)

IV. CONCLUSION

The model of spontaneous splitting advanced in this paper undoubtedly represents an oversimplification of the actual situation. Its main purpose is to illustrate the contention that it is not necessary to postulate the existence of a product structure at the outset, just as it is not necessary or desirable to presume the compactness of the internal space. All asymmetries of the vacuum should be explicable in terms of some physical mechanism. In this paper, some of the mathematical components of such a mechanism have been displayed.

Let P be a Ricci-flat compact Riemannian manifold on which is defined a gauge theory with gauge group J. Taking rank $J \geqslant m$, suppose that m gauge fields corresponding to a maximal torus of J develop vacuum expectation values. Then it can be shown[2] that if these fields are subject to the usual field equations and gauge fixing conditions, they must be constant. Thus P admits m constant vector fields, and so, according to the local de Rham decomposition theorem, P must be a local Riemannian product of a flat m-dimensional manifold M with a Ricci-flat manifold K which has a finite isometry group F. If this group F contains any freely acting cyclic subgroup, then it will always be possible to construct vacuum models in which the local splitting does not extend globally, which may have various unpleasant ramifications. The only certain way of eliminating this possibility is to factor out any cyclic free isometries; fortunately, there are in many cases good physical reasons for doing so. Thus finally we obtain a global decomposition, $P = M \times K$.

The problem of obtaining a transition from a local splitting to a global product can only be solved by introducing further global data. Fortunately, in superstring models such data are immediately available: the idea of factoring out the freely acting isometries of K was already suggested, for totally different reasons, in Ref. 1. The analysis here was motivated by the analogous observation[15] in the case of spontaneous compactification. In that case, the Freund–Rubin[3] mechanism ensures that the Ricci tensor of the internal space satisfies the correct conditions—but this is not sufficient to enforce compactification. An additional global condition, namely completeness, is needed, and again one must find a physical justification for imposing this condition. Fortunately, that can be done. The point in both these cases is that global asymmetries of the vacuum *can* arise spontaneously, provided that conditions on vacuum expectation values can be supplemented by appropriate physically motivated global conditions.

Finally, we should remark that recent work (see, for example, Ref. 16) has suggested that Calabi–Yau compactifications may not yield absolutely precise models of the superstring vacuum. It seems unlikely, however, that the discrepancy will be large enough to modify the qualitative picture at low energies. Thus it is probably still useful to discuss overall features of the vacuum, such as splitting, in the context provided by Ref. 1.

[1]P. Candelas, G. T. Horowitz, A. Strominger, and E. Witten, Nucl. Phys. B **258**, 46 (1985).

[2]B. T. McInnes, Class. Quantum Grav. **4**, 329 (1987).

[3]P. G. O. Freund and M. A. Rubin, Phys. Lett. B **97**, 233 (1980).

[4]K. Yano and S. Bochner, *Curvature and Betti Numbers* (Princeton U. P., Princeton, NJ, 1953).

[5]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I.

[6]J. Cheeger and D. G. Ebin, *Comparison Theorems in Riemannian Geometry* (North-Holland, Amsterdam, 1975).

[7]J. C. Graves, *The Conceptual Foundations of Contemporary Relativity Theory* (MIT, Cambridge, 1971).

[8]T. Eguchi, P. B. Gilkey, and A. J. Hanson, Phys. Rep. **66**, 213 (1980).

[9]G. R. Jensen, J. Differ. Geom. **8**, 599 (1973).

[10]Y. M. Cho, J. Math. Phys. **16**, 2029 (1975).

[11]E. Calabi, in *Algebraic Geometry and Topology: A Symposium in Honor of S. Lefschetz* (Princeton U. P., Princeton, NJ, 1957).

[12]K. Nomizu, Nagoya Math. J. **10**, 101 (1956).

[13]A. Strominger and E. Witten, Commun. Math. Phys. **101**, 341 (1985).

[14]S. T. Yau, in *Symposium on Anomalies, Geometry, Topology* (World Scientific, Singapore, 1985).

[15]B. T. McInnes, Class. Quantum Grav. **2**, 661 (1985).

[16]M. D. Freeman and C. N. Pope, Phys. Lett. B **174**, 48 (1986).

Canonical stochastic dynamical systems

Tetsuya Misawa
Department of Applied Mathematics and Physics, Kyoto University, Kyoto 606, Japan
Kunio Yasue
Research Institute for Informatics and Science, Notre Dame Seishin University, Okayama 700, Japan

(Received 30 May 1986; accepted for publication 1 July 1987)

A canonical stochastic dynamical system for time-symmetric semimartingales is formulated by the stochastic least action principle in a new stochastic calculus of variations. A certain class of stochastic dynamical systems gives a Hamiltonian formalism of Nelson's stochastic mechanics. In a manner analogous to classical mechanics, the notions of a stochastic Poisson bracket and canonical transformation are introduced to the stochastic dynamical systems. It is shown that the phase factor of the wave function plays the role of a generating function of the canonical transformation.

I. INTRODUCTION

Several years ago, one of the authors (K. Y.) proposed the notion of stochastic calculus of variations[1,2] within Nelson's mechanics.[3] Since that work there have been many attempts to develop the variational formulation of stochastic mechanics.[4–11]

On the variational principle in the original stochastic calculus of variation, Zambrini has developed a Hamiltonian formalism of stochastic mechanics.[5,6] However, the classical notions of canonical transformations were not extended straightforwardly to his stochastic mechanics. On the other hand, another author (T. M.) proposed a different Hamiltonian formalism from that of Zambrini.[12] Since the Hamiltonian formalism remained the canonical structure of classical mechanics, canonical transformation was easily formulated in it, but the formalism was not based on the variational principle.

In the present paper, we will develop a theory of canonical stochastic dynamical systems keeping the canonical structure of classical Hamiltonian mechanics on the basis of a new stochastic calculus of variations. The classical notions of Poisson bracket and canonical transformation[13] will be extended naturally to the new stochastic dynamical systems.

Section II is devoted to the exposition of a new stochastic calculus of variations, which is called the "symmetric stochastic calculus of variations (SSCV)." "Symmetric" means that this stochastic calculus employs a class of time-symmetric semimartingales.[5,6] We also touch upon the reformulation of Nelson's mechanics on the new stochastic least action principle in SSCV in Sec. II.

On the basis of the new stochastic principle, canonical stochastic dynamical systems are formulated for time-symmetric semimartingales in Sec. III. "Stochastic canonical equations" are first derived from the stochastic least action principle in phase space. A phase space-valued time-symmetric semimartingale is called a canonical stochastic dynamical system if the stochastic process satisfies the stochastic canonical equations. The connection between canonical stochastic dynamical systems and Nelson's mechanics is also investigated in this section.

In Sec. IV, a stochastic Poisson bracket is introduced and thereby a notion of the strongly conservative variable is defined in canonical stochastic dynamical systems.

In Sec. V, the canonical transformation is formulated in our canonical systems. It is shown that the phase factor of the wave function plays the role of a generating function of a canonical transformation.

Throughout this paper, we refer to Zambrini's review article[5] for the basic notions of stochastic calculus and stochastic mechanics.

II. SYMMETRIC STOCHASTIC CALCULUS OF VARIATIONS

In this section, in order to formulate canonical stochastic dynamical systems later on, we develop an idea of a new stochastic calculus of variations.

We start with a brief presentation of the class of stochastic processes which are taken into account in our stochastic calculus of variations.[5,6] Let (Ω,B,P) be a base probability space and x be a stochastic process in R^l, i.e., a continuous mapping $t\to x(t)$ from a time interval I into the Hilbert space $H=L^2((\Omega,P)\to R^l)$. We consider two filtrations indexed by I; B_t and $\widetilde{B}_t$ with $B_s\subset B_t$ and $\widetilde{B}_t\subset\widetilde{B}_s$ for $s\leqslant t$ to which $x(t)$ is adapted. By hypothesis, $x(t)$ is simultaneously a B_t semimartingale and a $\widetilde{B}_t$ semimartingale.[5,6] Moreover, the process $x(t)$ has the two mean velocities

$$Dx(t)=\lim_{h\to 0+} h^{-1}E\,[x(t+h)-x(t)|B_t] \tag{1}$$

and

$$D_*x(t)=\lim_{h\to 0+} h^{-1}E\,[x(t)-x(t-h)|\widetilde{B}_t], \tag{2}$$

where $E[\cdot|\beta]$ denotes the conditional expectation with respect to the σ-algebra β. We assume that these two limits exist in the Hilbert space H and the mappings $t\to Dx(t)$, $t\to D_*x(t)$ are continuous from I into H. Let us denote the class of stochastic processes of the above-mentioned type by K. The following symmetric integration by parts formula due to Zheng and Meyer[10] will be used for two processes $x(t)$ and $y(t)$ belonging to K,

 0022-2488/87/112569-05$02.50

$$E[x(b)\cdot y(b) - x(a)\cdot y(a)] = E\left[\int_a^b (y(t)\cdot D^\circ x(t) + x(t)\cdot D^\circ y(t))dt\right]. \quad (3)$$

In expression (3), $D^\circ x(t)$ denotes the "symmetric mean derivative" defined by

$$D^\circ x(t) \equiv \tfrac{1}{2}(D + D_*)x(t). \quad (4)$$

Now, we will formulate the new stochastic calculus of variations, which is called the symmetric stochastic calculus of variations (SSCV). For $L \in C^2(R^{2l+1} \to R^1)$, a Lagrangian, and for each process $x(t)$ in K, we define the action functional J by

$$J[x] = E\left[\int_a^b L(x(t), D^\circ x(t), t)dt\right], \quad (5)$$

where $E[\cdot]$ is the absolute expectation and $a,b \in I$, $a \leqslant b$. We denote by Δ the totality of processes $z(t) = z(x(t),t)$, where $z = z(x,t)$ is any smooth R^l-valued function vanishing identically for $t = a$ and b. We note that each process $x(t)$ in Δ also belongs to K.

The process $x(t)$ in K is called a stationary process of the functional J given by (5), if $\delta J[x](z)$, the first variation of the functional J in x on K, is equal to zero for any processes z in Δ. A computation with the Taylor expansion in the Lagrangian L and formula (4) are put together to show that

$$\delta J[x](z) = E\left[\int_a^b \left(D^\circ \frac{\partial L}{\partial D^\circ x(t)} - \frac{\partial L}{\partial x(t)}\right) z(t)dt\right]. \quad (6)$$

Here we have used $z(a) = z(b) = 0$. Therefore we see that a process $x(t)$ belonging to K is a stationary process if and only if for the process $x(t)$ the following equation holds:

$$D^\circ \frac{\partial L}{\partial D^\circ x(t)} - \frac{\partial L}{\partial x(t)} = 0. \quad (7)$$

We call Eq. (7) the stochastic Euler equation. This result describes the stochastic least action principle in SSCV.

Now, we will quickly touch upon the reformulation of Nelson's stochastic mechanics[3] on our stochastic least action principle. Some results in the following will be used to associate canonical stochastic dynamical systems with Nelson's mechanics later on. Let us consider the diffusion processes $x(t)$ belonging to K which are governed by the stochastic differential equation and the reversed equation[5]:

$$dx(t) = b(x(t),t)dt + \{\hbar/(2m)\}^{1/2} dw(t), \quad (8)$$

$$dx(t) = b_*(x(t),t)dt + \{\hbar/(2m)\}^{1/2} dw_*(t), \quad (9)$$

where b and b_* are certain vector-valued smooth functions, $\hbar$ is Planck's constant divided by 2π, and m is a mass of a particle. In (8) and (9), $w(t)$ is a standard R^l-valued Wiener process and $w_*(t)$ has the same properties as $w(t)$ except that the increments $w_*(t) - w_*(s)$ are independent of $x(\tau)$ for $\tau \geqslant t > s$. We assume that $x(t)$ has a probability density function $\rho(x,t)$. For this process we have $Dx(t) = b(x(t),t)$ and $D_*x(t) = b_*(x(t),t)$. According to Nelson,[3] these functions b and b_* are connected with the probability density ρ by the following equations:

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(v\cdot\rho) = 0, \quad u = \left\{\frac{\hbar}{(2m)}\right\} \operatorname{grad} \ln \rho, \quad (10)$$

where v, the "current velocity," and u, the "osmotic velocity," are vector-valued functions defined by $v = \frac{1}{2}(b + b_*)$ and $u = \frac{1}{2}(b - b_*)$, respectively.

Now, we assume that the diffusion process $x(t)$ mentioned above is an extremal of the functional J with the following Lagrangian:

$$L(x, D^\circ x, t) = (m/2)|D^\circ x|^2 - U(x,t), \quad (11)$$

where

$$U(x,t) = V(x,t) + \frac{\hbar^2}{2m}\left(\frac{\Delta\{\rho(x,t)\}^{1/2}}{\{\rho(x,t)\}^{1/2}}\right). \quad (12)$$

In (12), V is a given potential function. Then inserting Eq. (11) with (12) into Eq. (7), we obtain

$$\left(\frac{\partial}{\partial t} + v\cdot\nabla\right)v(x(t),t) = -\frac{1}{m}\operatorname{grad} V(x(t),t) + \left(u\cdot\nabla + \frac{\hbar}{2m}\Delta\right)u(x(t),t). \quad (13)$$

Here we have used the second equation of Eq. (10) and the following equation[5]:

$$D^\circ v(x(t),t) = \left[\frac{\partial}{\partial t} + v\cdot\nabla\right]v(x(t),t). \quad (14)$$

Equation (13) is just the same consequence as that of "Newton's equation of motion" in Nelson's mechanics. Therefore in a manner analogous to Nelson's mechanics we can determine v, u, and ρ (and hence b and b_*) from Eqs. (10) and (13), so that the diffusion process $x(t)$ is determined.[3] Furthermore, one can show that the diffusion process $x(t)$ together with the Lagrangian (11) corresponds to a solution of a Schrödinger equation. Let us consider the wave function defined by

$$\Psi(x,t) = \{\rho(x,t)\}^{1/2} \exp\{iS(x,t)\}, \quad (15)$$

where $S(x,t)$ is such that

$$v(x,t) = (\hbar/m) \operatorname{grad} S(x,t). \quad (16)$$

Then the function $\Psi(x,t)$ satisfies the Schrödinger equation with the potential function $V(x,t)$,[3]

$$i\frac{\partial\Psi}{\partial t} = -\frac{\hbar}{2m}\Delta\Psi + \frac{1}{\hbar}V\Psi. \quad (17)$$

From the first equation of Eq. (10), Eq. (13) and the definition of S, ρ and S prove

$$\frac{\partial\rho}{\partial t} + \frac{1}{m}\operatorname{div}(\rho\cdot\operatorname{grad} S) = 0, \quad (18)$$

and

$$\frac{\partial S}{\partial t} + \frac{1}{2m}(\operatorname{grad} S)^2 + V(x,t) + \frac{\hbar^2}{2m}\left(\frac{\Delta\{\rho(x,t)\}^{1/2}}{\{\rho(x,t)\}^{1/2}}\right) = 0. \quad (19)$$

Hence $\rho = |\Psi|^2$ gives the probability density of a particle in the position space. Thus Nelson's stochastic mechanics is now reformulated on the stochastic least action principle.

III. A CANONICAL STOCHASTIC DYNAMICAL SYSTEM

In this section, we associate the time-symmetric semimartingales in the class K with canonical stochastic dynamical sytems. At the first stage, we derive stochastic canonical equations from our new stochastic least action principle in Sec. II.

Let us introduce the canonical momentum process $p(t)$ by

$$p(t) = \frac{\partial L}{\partial D^\circ x(t)} (x(t), D^\circ x(t), t) \tag{20}$$

and the Hamiltonian

$$H(x(t),p(t),t) = p(t)\cdot D^\circ x(t) - L(x(t),D^\circ x(t),t). \tag{21}$$

We assume here that two stochastic processes $x(t)$ and $p(t)$ belong to the class K, respectively. Then, we may regard a pair of processes $(x(t),p(t))$ as a phase space-valued (i.e., $R^l \times R^l$-valued) time-symmetric semimartingale.

Now, we consider the functional for a pair $(x(t),p(t))$,

$$I[x,p] = E\left[\int_a^b \{p(t)\cdot D^\circ x(t) - H(x(t),p(t),t)\}dt\right]. \tag{22}$$

Suppose that $(x(t),p(t))$ is an extremal of this functional under the stochastic variation

$$(x(t),p(t)) + (z_1(x(t),p(t),t), z_2(x(t),p(t),t)),$$

where z_1 and z_2 are any R^l-valued smooth functions vanishing identically for $t=a$ and b. This means the stochastic least action principle in phase space. As in classical particle dynamics, the stochastic least action principle in SSCV applied to Eq. (22) yields stochastic canonical equations

$$D^\circ x(t) = \frac{\partial}{\partial p(t)} H(x(t),p(t),t), \tag{23}$$

$$D^\circ p(t) = -\frac{\partial}{\partial x(t)} H(x(t),p(t),t). \tag{24}$$

We remark that these equations correspond to the basic equations in the first author's mechanics.[12] We call a pair of stochastic processes $(x(t),p(t))$ belonging to K a canonical stochastic dynamical system if $(x(t),p(t))$ satisfies stochastic canonical equations (23) and (24).

Now, we will mention here the relation of our canonical stochastic dynamical systems to Nelson's stochastic mechanics. Suppose that $(x(t),p(t))$ is a canonical stochastic dynamical system, which satisfies the following conditions.

(i) $x(t)$ is a diffusion process governed by the stochastic differential equation (8) and the reversed equation (9).

(ii) The Hamiltonian is given by

$$H(x,p,t) = \frac{1}{2m}|p|^2 + V(x,t) + \frac{\hbar^2}{2m}\left(\frac{\Delta\{\rho(x,t)\}^{1/2}}{\{\rho(x,t)\}^{1/2}}\right). \tag{25}$$

Then we can show that the process $x(t)$ satisfies Eq. (13) which is one of the fundamental equations in stochastic mechanics. Indeed, for the Hamiltonian (25), canonical equations (23) and (24) turn into

$$\begin{aligned} p(t) &= mD^\circ x(t) \\ &= mv(x(t),t), \end{aligned} \tag{26}$$

and

$$D^\circ p(t) = -\frac{1}{m}\operatorname{grad} V(x,t) - \frac{\hbar^2}{2m}\operatorname{grad}\left(\frac{\Delta\{\rho(x,t)\}^{1/2}}{\{\rho(x,t)\}^{1/2}}\right), \tag{27}$$

where v is the "current velocity" defined in Sec. II. Inserting Eq. (26) into Eq. (27), and using the second equation of Eqs. (10) and (14), we obtain

$$\left(\frac{\partial}{\partial t} + v\cdot\nabla\right)v(x(t),t) = -\frac{1}{m}\operatorname{grad} V(x(t),t) + \left(u\cdot\nabla + \frac{\hbar}{2m}\Delta\right)u(x(t),t)$$

as we wanted. Therefore in the same manner as was mentioned in the end of Sec. II, we can determine the diffusion process $x(t)$ from this equation together with Eq. (10), and further show that the process is corresponding to a solution of the Schrödinger equation. Then, the process $p(t)$ is determined from $x(t)$ through Eq. (26). From Eqs. (16) and (26), we see that $p(t)$ is connected with the phase factor of the wave function (15) through the following equation:

$$p(t) = \hbar \operatorname{grad} S(x(t),t). \tag{28}$$

Thus we can associate the canonical stochastic dynamical system $(x(t),p(t))$ with the Schrödinger equation (17).

The above-mentioned results tell us that a certain class of canonical stochastic dynamical systems gives a Hamiltonian formalism of Nelson's stochastic mecahanics. Therefore we may regard our canonical dynamical systems as an extension of Nelson's mechanics.

IV. STOCHASTIC POISSON BRACKET

In this section, we define the stochastic Poisson bracket and thereby set up the notion of a strongly conservative variable in canonical stochastic dynamical systems. In what follows, a pair of processes $(x(t),p(t))$ is a canonical stochastic dynamical system with a Hamiltonian $H(x,p,t)$.

Let us consider a dynamical variable in canonical stochastic dynamical systems $g(x(t),p(t),t)$, where g is a smooth function. We compute a stochastic differential of $g(x(t),p(t),t)$,[5]

$$\begin{aligned} dg(x(t),p(t),t) &= \frac{\partial g}{\partial t}(x(t),p(t),t)dt + \frac{\partial g}{\partial x(t)}(x(t),p(t),t)\circ dx(t) \\ &\quad + \frac{\partial g}{\partial p(t)}(x(t),p(t),t)\circ dp(t), \end{aligned} \tag{29}$$

where $\circ$ denotes the symmetric stochastic differential of Fisk and Stratonovich. Taking the expectation, we have

$$\begin{aligned} E[dg(x(t),p(t),t)] &= E\left[\left\{\frac{\partial g}{\partial t}(x(t),p(t),t) + \frac{\partial g}{\partial x(t)}(x(t),p(t),t)\cdot D^\circ x(t)\right.\right. \\ &\quad \left.\left. + \frac{\partial g}{\partial p(t)}(x(t),p(t),t)\cdot D^\circ p(t)\right\}dt\right]. \end{aligned} \tag{30}$$

Here we have used the following formula[5]:

$$E[y(t)\circ dx(t)] = E[y(t)\cdot D^\circ x(t)dt]. \tag{31}$$

By the canonical equations (23) and (24), we have

$$\frac{d}{dt}E\,[g(x(t),p(t),t)]$$
$$= E\left[\frac{\partial g}{\partial t}(x(t),p(t),t) + \frac{\partial g}{\partial x(t)}(x(t),p(t),t)\cdot\frac{\partial H}{\partial p(t)}(x(t),p(t),t) - \frac{\partial g}{\partial p(t)}(x(t),p(t),t)\cdot\frac{\partial H}{\partial x(t)}(x(t),p(t),t)\right]. \quad (32)$$

This expression tells us how the expectation value of the dynamical variable $g(x(t),p(t),t)$ changes in canonical stochastic dynamical systems.

Now, we introduce here the notion of a stochastic Poisson bracket. Let $f = f(x(t),p(t),t)$ and $g = g(x(t),p(t),t)$ be two dynamical variables in canonical stochastic dynamical systems. Then, a stochastic Poisson bracket of those dynamical variables is defined by

$$\{f(x(t),p(t),t),g(x(t),p(t),t)\} \equiv \left(\frac{\partial f}{\partial p(t)}\cdot\frac{\partial g}{\partial x(t)} - \frac{\partial f}{\partial x(t)}\cdot\frac{\partial g}{\partial p(t)}\right)(x(t),p(t),t). \quad (33)$$

It has the same form as the usual Poisson bracket in classical mechanics[13] due to the symmetric stochastic calculus. The stochastic Poisson bracket satisfies the following properties:

$$\{f(x(t),p(t),t),x^i(t)\} = \frac{\partial f}{\partial p_i(t)}(x(t),p(t),t), \quad (34)$$

$$\{f(x(t),p(t),t),p_i(t)\} = -\frac{\partial f}{\partial x^i(t)}(x(t),p(t),t), \quad (35)$$

$$\{x^i(t),x^j(t)\} = \{p_i(t),p_j(t)\} = 0, \quad (36)$$

$$\{p_i(t),x^j(t)\} = \delta_i^j, \quad (37)$$

for $1 \leqslant i,j \leqslant l$. Moreover, the Jacobi identity

$$\{f,\{g,h\}\} + \{g,\{h,f\}\} + \{h,\{f,g\}\} = 0 \quad (38)$$

holds, where $h = h(x(t),p(t),t)$ is another dynamical variable.

In terms of the stochastic Poisson bracket, Eq. (32) becomes

$$\frac{d}{dt}E\,[g(x(t),p(t),t)] = E\left[\frac{\partial g}{\partial t}(x(t),p(t),t) + \{H(x(t),p(t),t),g(x(t),p(t),t)\}\right]. \quad (39)$$

Now, we coin a new notion of a strongly conservative variable. A dynamical variable $h(x(t),p(t))$ that does not depend on time explicitly is said to be strongly conservative if its stochastic Poisson bracket with the Hamiltonian $H(x(t),p(t),t)$ vanishes, that is,

$$\{H(x(t),p(t),t),h(x(t),p(t))\} = 0. \quad (40)$$

Expectation of such a strongly conservative variable is a constant of motion, namely,

$$\frac{d}{dt}E\,[h(x(t),p(t))] = 0. \quad (41)$$

It is worthwhile to notice here that two known strongly conservative variables $h_1(x(t),p(t))$ and $h_2(x(t),p(t))$ generate another one by their stochastic Poisson bracket

$$h_3(x(t),p(t)) = \{h_1(x(t),p(t)),h_2(x(t),p(t))\}. \quad (42)$$

This fact corresponding to Poisson's theorem in classical mechanics[13] is a direct consequence of the Jacobi identity (38). Using this Poisson's theorem successively, a family of strongly conservative variables in canonical stochastic dynamical systems will be obtained provided that the first two are known.

V. CANONICAL TRANSFORMATIONS

In a manner similar to classical mechanics,[13] we will introduce canonical transformations in canonical stochastic dynamical systems. Let $X = X(x,p,t)$ and $P = P(x,p,t)$ be two smooth functions. For each canonical stochastic dynamical system $(x(t),p(t))$, we define another one $(X(t),P(t))$ by

$$X(t) = X(x(t),p(t),t), \quad (43)$$

$$P(t) = P(x(t),p(t),t). \quad (44)$$

Such a transformation in phase space is said to be a canonical transformation if $(X(t),P(t))$ satisfies the stochastic canonical equations (23) and (24),

$$D\circ X(t) = \frac{\partial \widetilde{H}}{\partial P(t)}(X(t),P(t),t), \quad (45)$$

$$D\circ P(t) = -\frac{\partial \widetilde{H}}{\partial X(t)}(X(t),P(t),t). \quad (46)$$

Here, $\widetilde{H} = \widetilde{H}(X,P,t)$ is a certain Hamiltonian.

The canonical stochastic dynamical system $(X(t),P(t))$ extremizes, then, the action integral

$$\widetilde{I}[X,P] = E\left[\int_a^b (P(t)\cdot D\circ X(t) - \widetilde{H}(X(t),P(t),t))dt\right]. \quad (47)$$

Since $(x(t),p(t))$ extremizes the action integral $I[x,p]$ given by Eq. (22), the difference between the two stochastic integrals

$$\int_a^b (p(t)\cdot D\circ x(t) - H(x(t),p(t),t))dt, \quad (48)$$

and

$$\int_a^b (P(t)\cdot D\circ X(t) - \widetilde{H}(X(t),P(t),t))dt, \quad (49)$$

must be of the form

$$\int_a^b dF(t) \quad (50)$$

for a stochastic process $F(t)$. Here, a new degree of freedom arises from the choice of $F(t)$.

For example, we can choose

$$F(t) = F(x(t),X(t),t), \quad (51)$$

where $F = F(x,X,t)$ is a smooth function of x, its image X under the canonical transformation, and time. Such a function F as Eq. (51) is called a generating function of the canonical transformation (43) and (44). Since

$$p(t)\circ dx(t) - H(x(t),p(t),t)dt - P(t)\circ dX(t) + \widetilde{H}(X(t),P(t),t)dt$$
$$= dF(x(t),X(t),t) = \frac{\partial F}{\partial x(t)}(x(t),X(t),t)\circ dx(t) + \frac{\partial F}{\partial X(t)}(x(t),X(t),t)\circ dX(t) + \frac{\partial F}{\partial t}(x(t),X(t),t)dt,$$

we find

$$p(t) = \frac{\partial F}{\partial x(t)}(x(t),X(t),t), \tag{52}$$

$$P(t) = \frac{\partial F}{\partial X(t)}(x(t),X(t),t), \tag{53}$$

and

$$\widetilde{H}(X(t),P(t),t) = H(x(t),p(t),t) + \frac{\partial F}{\partial t}(x(t),X(t),t). \tag{54}$$

Thus a generating function defines a canonical transformation of the stochastic processes in canonical stochastic dynamical systems.

Let us consider a generating function of the type

$$S(t) = S(x(t),t), \tag{55}$$

where $S = S(x,t)$ is a smooth function. Suppose the Hamiltonian $\widetilde{H}$ vanishes identically by the canonical transformation generated by S. Then Eqs. (52)–(54) yield

$$p(t) = \frac{\partial S}{\partial x(t)}(x(t),t), \tag{56}$$

$$P(t) = 0, \tag{57}$$

$$-\frac{\partial S}{\partial t}(x(t),t) = H(x(t),p(t),t), \tag{58}$$

respectively. Equations (56) and (58) assert that the generating function S should be subject to

$$-\frac{\partial S}{\partial t}(x,t) = H\left(x,\frac{\partial S}{\partial x}(x,t),t\right). \tag{59}$$

This equation extends the Hamilton–Jacobi equation in classical mechanics[13] to the canonical stochastic dynamical systems. We call it the stochastic Hamilton–Jacobi equation. If the Hamiltonian H is given by Eq. (25), then Eq. (59) turns into

$$-\frac{\partial S}{\partial t} = \frac{1}{2m}\left(\frac{\partial S}{\partial x}\right)^2 + V(x,t) + \frac{\hbar^2}{2m}\left(\frac{\Delta\{\rho(x,t)\}^{1/2}}{\{\rho(x,t)\}^{1/2}}\right). \tag{60}$$

This coincides with Eq. (19). Therefore we see that the phase factor of the wave function in Nelson's mechanics is nothing else but a generating function of a canonical transformation which makes the stochastic canonical equation much simpler.

[1]K. Yasue, J. Math. Phys. **22**, 1010 (1981).
[2]K. Yasue, J. Funct. Anal. **41**, 327 (1981).
[3]E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U.P., Princeton, 1964).
[4]K. Yasue, J. Funct. Anal. **51**, 133 (1983).
[5]J.-C. Zambrini, Int. J. Theor. Phys. **24**, 277 (1985).
[6]J.-C. Zambrini, J. Math. Phys. **25**, 1314 (1984).
[7]J.-C. Zambrini and K. Yasue, Ann. Phys. (NY) **143**, 54 (1982).
[8]J.-C. Zambrini and K. Yasue, Phys. Rev. Lett. **52**, 2107 (1984).
[9]J.-C. Zambrini, Phys. Rev. A **33**, 1532 (1986).
[10]W. A. Zheng and P. A. Meyer, *Lecture Notes in Mathematics*, Vol. 18, Seminaire de Probabilites (Springer, Berlin, 1984), p. 233
[11]E. Nelson, *Quantum Fluctuations* (Princeton U.P., Princeton, NJ, 1985).
[12]T. Misawa, Nuovo Cimento B **91**, 1 (1986).
[13]H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, MA, 1980), 2nd ed.

Deformation of algebras and solution of self-duality equation

A. N. Leznov and M. A. Mukhtarov
Institute for High Energy Physics, P.O. Box 142284, Protvino, Serpukhov, Moscow, USSR

(Received 26 February 1987; accepted for publication 20 May 1987)

The solutions of the self-duality equation depending on a set of functions of three independent variables are constructed in an explicit way. The obtained solutions are shown to be connected with two-dimensional systems determined by the operators of the Lax pair.

I. INTRODUCTION

From the experience of integrating two-dimensional systems the importance of the deformation operation of internal symmetry algebras of the equations under consideration has been revealed. When one applies the Inonü–Wigner "contraction" operation, nonlinearities in dynamical systems disappear and systems transform into linear ones. However, the internal symmetry algebras of linear and nonlinear systems are the same, and thus makes it possible to connect them by Bäcklund transformation. Then the solutions of nonlinear systems are obtained as functions and their derivatives from linear equation solutions.

In the present paper we propose to carry out an analogous procedure for obtaining a class of solutions for four-dimensional self-duality equations that depend on a definite number of arbitrary functions of three independent variables. In the case of an arbitrary semisimple group the self-duality equations are a system of n second-order equations (n is the group dimension, r is its rank). The general solution of such a system must depend on $2n$ arbitrary functions of three independent variables and, if such a solution is obtained, one can expect the Cauchy problem to be solved. The technique of the present paper gives us a way to construct solutions as perturbative expansion series dependent on $(n+r)$ arbitrary functions being less than $2n$ and in this sense the constructed solutions are not a complete solution of the problem and we are not able to classify them yet. The recurrent procedure makes it possible to calculate in an explicit form every order of a perturbative expansion series by quadratures and algebraic operations and, what is more, the result of the summation of the series is equal to the solution of two-dimensional equations defined by the Lax operators under the condition that the values of arbitrary functions, when the parameter is equal to zero, are in accordance with certain solutions of four-dimensional self-duality equations when the parameter of interaction vanishes. At each intermediate stage the integration over real four-dimensional space does not take place and the whole apparatus of perturbation theory corresponds to single-multiple integrals.

We have not managed to find the physical interpretation of the integration parameter and the connection of the constructed solution with the known monopole and instanton configurations.

In Sec. II self-duality equations are given in the form we will use. The information about the Inonü–Wigner contraction operation is given and it is demonstrated that in this case, for definiteness, of semisimple algebras, the perturbation theory is equal to the integration of a system of equations with an infinite-dimensional solvable algebra which is a non-negative part of the infinite-dimensional algebra of a moderate rise.

In Sec. III recurrent relations connecting different orders of perturbative expansion series are obtained and with the help of them the formula for the nth-order term is calculated.

In Sec. IV it is shown that the summation of a perturbative expansion series is equal to the solution of a two-dimensional system of equations defined by operators of the Lax pair when the interaction parameter is equal to zero.

The final notes and conclusions are collected in Sec. V.

II. ALGEBRAIC PROPERTIES OF A PERTURBATIVE EXPANSION SERIES

The system of equations for self-dual configuration of Yang–Mills fields can be represented as one equation on the parameter of the group's element G (for definiteness of the semisimple group):

$$(G_{\bar z}G^{-1})_z + (G_{\bar y}G^{-1})_{\bar y} = 0,$$
$$(G^{-1}G_z)_{\bar z} + (G^{-1}G_y)_{\bar y} = 0, \qquad (1)$$

where z, $\bar z$, y, and $\bar y$ are independent variables, $G_y = \partial G/\partial y$. Equation (1) may be partially solved:

$$G_{\bar z}G^{-1} = f_y, \quad G_{\bar y}G^{-1} = -f_z,$$

where the element f belongs to the corresponding group algebra. From (1) and the definition of f there follows an equaton

$$\Box f + [f_z, f_y] = 0, \qquad (2)$$

where

$$\Box = \frac{\partial^2}{\partial y\,\partial \bar y} + \frac{\partial^2}{\partial z\,\partial \bar z}.$$

We shall call Eq. (2) a self-duality equation. Equation (2), contrary to (1), has the Lagrangian

$$L = \tfrac{1}{2} sp(f_y f_{\bar y} + f_z f_{\bar z}) + \tfrac{1}{3} sp(f\,[f_y f_z]). \qquad (3)$$

One can be convinced of it by constructing Euler's equations. For symmetrical properties of Eq. (2) refer to Ref. 1. The coupling constant in gauge theories is always introduced as a coefficient at the terms quadratic on fields and from this point of view the introduction of the interaction constant corresponds to adding the factor g^2 at the term $[f_y, f_z]$ in Eq. (2). From our experience working with two-dimensional systems[2] we introduce the interaction constant in a different way as the parameter of group Inonü–Wigner deformation in the corresponding semisimple algebra, i.e.,

 0022-2488/87/112574-05$02.50

we will consider that the element f belongs not to a semisimple algebra but to an algebra following from it by Inonü–Wigner deformation. Technically it means that the commutation relations between simple roots and Cartan elements of the semisimple algebra

$$[X_\alpha^+,X_\beta^-]=\delta_{\alpha\beta}h_\alpha,\quad [h_\alpha,X_\beta^\pm]=\pm K_{\beta\alpha}X_\beta^\pm \tag{4a}$$

transform into

$$[X_\alpha'^+,X_\beta'^-]=g^2\delta_{\alpha\beta}h'_\beta,\quad [h'_\alpha X_\beta'^\pm]=\pm K_{\beta\alpha}X_\beta'^\pm. \tag{4b}$$

It is easy to note that when the parameter is equal to zero the semisimple algebra transforms into a solvable one but the structure of its positive and negative spaces is the same as in a semisimple algebra. Incidentally the linear equations correspond to the vanishing value of the interaction constant and the integration of such equations is trivial. The internal symmetry algebra is not semisimple but solvable, defined by relations (4a) and (4b) when $g=0$. Then the algebras of the initial nonlinear system and of the linear system following from it by the contraction operation are the same and this makes it possible to connect them through a Bäcklund transformation and integrate them in an explicit form. At present we do not have independent methods for the reconstruction of the internal symmetry algebra from Eq. (2) in the case of four independent variables (compare with Ref. 3) and therefore such an introduction of an interaction parameter is an *a priori* assumption called for by the results of the present work.

The contraction operation is equal to multiplication of the compound root X_α on g^{n_α}, where n_α is its height (the number of simple roots from which it is formed). The generators of the space with zero graduate index are Cartan elements. The terms $gX_\alpha^\pm$ are the elements of the space with graduate index unity. All pair commutations $[gX_\alpha^\pm,gX_\beta^\pm]$ and the Cartan elements g^2h_α are the elements of the space with the graduate index 2, that is, the selection of corresponding-order terms from system (2) is equivalent to considering that the element f belongs to the infinite-dimensional solvable algebra, whose structure of graduate subspaces is described above.

Let us consider in detail the algebra $A_1(\mathrm{SL}(2,R))$. Substituting the expansion $f=\alpha X^+ +\tau h+\beta X^-$ into Eq. (2) and using the commutation relations

$$[X^+,X^-]=g^2h,\quad [h,X^\pm]=\pm 2X^\pm,$$

we obtain

$$\begin{aligned}\Box\alpha X^+ &+\Box\tau h+\Box\beta X^-\\ &=2(\tau_y\alpha_z-\tau_z\alpha_y)X^+ +g^2(\alpha_y\beta_z-\alpha_z\beta_y)h\\ &\quad-2(\tau_y\beta_z-\tau_z\beta_y)X^-.\end{aligned}$$

Equating the coefficients at the linearly independent generators of the algebra we get the system of equations

$$\begin{aligned}&\Box\alpha=2(\tau_y\alpha_z-\tau_z\alpha_y),\quad \Box\beta=-2(\tau_y\beta_z-\tau_z\beta_y),\\ &\Box\tau=g^2(\alpha_y\beta_z-\alpha_z\beta_y).\end{aligned} \tag{5}$$

Expanding α, β, τ into a series for the constant g^2,

$$\alpha=\sum_{n=0}^{\infty}\alpha^n g^{2n},\quad \beta=\sum_{n=0}^{\infty}\beta^n g^{2n},\quad \tau=\sum_{n=0}^{\infty}\tau^n g^{2n},$$

and substituting into system (5) we obtain for the nth order of perturbation theory the following system:

$$\begin{aligned}&\Box\tau^n=\sum_{k=0}^{n-1}(\alpha_y^k\beta_z^{n-k-1}-\alpha_z^k\beta_y^{n-k-1}),\\ &\Box\alpha^n=2\sum_{k=0}^{n-1}(\tau_y^k\alpha_z^{n-k}-\tau_z^k\alpha_y^{n-k}),\\ &\Box\beta^n=-2\sum_{k=0}^{n}(\tau_y^k\beta_z^{n-k}-\tau_z^k\beta_y^{n-k}).\end{aligned} \tag{6}$$

The expansion $f=\alpha X^+ +\tau h+\beta X^-$ takes the form

$$f=\sum_{n=0}^{\infty}(\alpha^n g^{2n}X^+ +\tau^n g^{2n}h+\beta^n g^{2n}X^-).$$

Let us introduce the notation

$$X_k^\pm=g^{2k}X^\pm,\quad H_s=g^{2s}h.$$

Then the element f becomes expanded over the infinite-dimensional algebra with the commutation relations on generators

$$\begin{aligned}&[X_i^+,X_j^-]=g^{2(i+j)}h=H_{i+j},\\ &[X_i^\pm,H_s]=\mp X_{s+i}^\pm\end{aligned}$$

and this algebra is equal to the solvable part of the infinite-dimensional algebra of Kac–Moody[4] $\tilde{A}_2$.

III. THE EXPLICIT EXPRESSION OF THE *n*TH-ORDER TERMS IN THE CASE OF SOLVABLE ALGEBRAS

We will consider the algebra of Eq. (2) as being finite solvable or infinite dimensional. The basic vectors of the zero invariant subspace will be denoted by h_i and their corresponding functions by τ_i, that is, the element of the zero invariant subspace H is identically equal to $H=\Sigma_{i=1}^r h_i\tau_i$. From (2) it follows that $\Box H=0$ or

$$\Box\tau_i=0, \tag{7}$$

i.e., τ_i is the solution of the Dalamber equation, whose general solution depends on two arbitrary functions of three independent variables (see Appendix A).

For the functions of the first invariant subspace X^1, equal to $X^1=\Sigma_{\alpha=1}^r X_\alpha^+a^\alpha$, we have from (2)

$$\begin{aligned}&\Box X^1+[H_zX_y^1]+[X_z^1H_y]=0,\\ &\Box a^\alpha=\rho_y^\alpha a_z^\alpha-\rho_z^\alpha a_y^\alpha,\end{aligned} \tag{8}$$

where the ρ^α are defined by the relation

$$[HX_\alpha^+]=\rho^\alpha X_\alpha^+.$$

For the often occurring combinations of the functions $(\phi_yf_z-\phi_zf_y)$ we introduce the notation $\{\phi,f\}_{y,z}$. Equation (8) possesses the Bäcklund transformation of the following form:

$$a_{\bar z}^\alpha-\rho_y^\alpha a^\alpha=b_y^\alpha,\quad a_{\bar y}^\alpha+\rho_z^\alpha a^\alpha=-b_z^\alpha. \tag{9}$$

From these relations it follows that the functions b^α satisfy the same equation [Eq. (8)] when ρ^α satisfies the equation

$$\Box\rho^\alpha=0.$$

We will make the assumption that $b^\alpha=\lambda a^\alpha$. Then one rewrites system (9) in the form

$$a_{\bar z}^\alpha=\lambda a_y^\alpha+\rho_y^\alpha a^\alpha,\quad a_{\bar y}^\alpha=-\lambda a_z^\alpha-\rho_z^\alpha a^\alpha. \tag{10}$$

Solving (10) we obtain a particular solution for (8) of the following form:

$$a^\alpha = A^\alpha(y+\lambda\bar{z}, z-\alpha\bar{y},\lambda)\exp B^\alpha, \tag{11}$$

where

$$B^\alpha = \frac{1}{2}\left(\frac{\partial/\partial y}{\partial/\partial\bar{z}-\lambda(\partial/\partial y)} - \frac{\partial/\partial z}{\partial/\partial\bar{y}+\lambda(\partial/\partial z)}\right)\rho^\alpha.$$

Due to the fact that λ is an arbitrary parameter and the initial equation is linear for a^α, the sum of solutions is again a solution and therefore $\int a^\alpha(\lambda)d\lambda$ will be the solution of Eq. (8), too. As is seen from the explicit relation (11) the solution depends on one arbitrary function of three arguments, but the general solution must depend on two arbitrary functions, thus the constructed solution is a particular one. The equation for the function of the second invariant subspace X^2,

$$X^2 = \sum_{\alpha<\beta} a^{\alpha\beta}X^+_{\alpha\beta}, \quad X^+_{\alpha\beta} = [X^+_\alpha, X^+_\beta],$$

has the form

$$\Box X^2 + [H_z X^2_y] + [H^1_y X^2_z] + [X^1_z X^1_y] = 0$$

or

$$\Box a^{\alpha\beta} + \{\rho^\alpha + \rho^\beta, a^{\alpha\beta}\}_{zy} + \{a^\alpha, a^\beta\}_{zy} = 0. \tag{12}$$

Equation (12) is a nonhomogeneous one, whose nonhomogeneities are given by the constructed functions a^α. One of the particular solutions of this equation has the following form:

$$\beta^{\alpha\beta} = \frac{1}{2}\int a^{\alpha\beta}(\lambda)d\lambda,$$

where

$$a^{\alpha\beta} = [a^\alpha(\lambda), a^\beta(\lambda)]^1.$$

By the commutation of two functions f, g we obtain the following expression:

$$[f(\lambda),g(\lambda)]^1 = f(\lambda)\int\frac{g(\lambda')d\lambda'}{\lambda-\lambda'} - g(\lambda)\int\frac{f(\lambda')d\lambda'}{\lambda-\lambda'}.$$

Using relations (10) and carrying out a number of identical algebraic transformations we get that $\frac{1}{2}\int[a^\alpha(\lambda)a^\beta(\lambda)]^1\,d\lambda$ satisfies (12). Adding to the obtained solution the solution of a homogeneous equation according to scheme (10) we get a particular solution of (12) dependent on one more function of three variables. The general scheme of the reduction proved in Appendix B is the following: for the function of the nth invariant subspace we have the equation

$$\Box X^n + [H_z X^n_y] + [X^n_z H_y] + \sum_{k=1}^{n-1}[X^k_z, X^{n-k-1}_y] = 0.$$

The assertion of reduction consists of the fact that the particular solution of this equation

$$B^{\alpha_1,\alpha_2,\ldots,\alpha_n} = \frac{1}{n!}A^{\alpha_1,\alpha_2,\ldots,\alpha_n} = \frac{1}{n!}\int a^{\alpha_1,\alpha_2,\ldots,\alpha_n}(\lambda)d\lambda$$

is defined by the use of previous reduction stages by the formulas

$$a^{\alpha_1,\alpha_2,\ldots,\alpha_n} = \sum_{s=0}^{n-1} C^{s-1}_{n-2}[a^{\alpha_1,\ldots,\alpha_s}, a^{\alpha_{s+1},\ldots,\alpha_n}]^1. \tag{13}$$

For the function $a^{\alpha_1,\alpha_2,\ldots,\alpha_n}$ the following relation takes place:

$$\begin{aligned} a^{\alpha_1,\ldots,\alpha_n}_{\bar{z}} &= \lambda a^{\alpha_1,\ldots,\alpha_n}_y + \left(\sum_{i=1}^n \rho^{\alpha_i}\right)a^{\alpha_1,\ldots,\alpha_n} \\ &\quad + \sum_{i=0}^{n-2}(-C^i_{n-1}a^{\alpha_1,\ldots,\alpha_{i+1}}A^{\alpha_{i+2},\ldots,\alpha_n}_y \\ &\quad + C^{i+1}_{n-1}a^{\alpha_{i+2},\ldots,\alpha_n}A^{\alpha_1,\ldots,\alpha_{i+1}}_y), \\ a^{\alpha_1,\ldots,\alpha_n}_{\bar{y}} &= -\lambda a^{\alpha_1,\ldots,\alpha_n}_z - \left(\sum_{i=1}^n \rho^{\alpha_i}\right)a^{\alpha_1,\ldots,\alpha_n} \\ &\quad + \sum_{i=0}^{n-2}(C^i_{n-1}a^{\alpha_1,\ldots,\alpha_{i+1}}A^{\alpha_{i+2},\ldots,\alpha_n} \\ &\quad - C^{i+1}_{n-1}a^{\alpha_{i+2},\ldots,\alpha_n}A^{\alpha_1,\ldots,\alpha_{i+1}}). \end{aligned} \tag{14}$$

Differentiating the first relation with respect to z, the second with respect to y, and taking the sum, one comes to the relation

$$\begin{aligned} \Box a^{\alpha_1,\ldots,\alpha_n} &= \left\{\sum_{i=1}^n \rho^{\alpha_i}, a^{\alpha_1,\ldots,\alpha_n}\right\}_{yz} \\ &\quad + \sum_{i=0}^{n-2}(-C^i_{n-1}\{A^{\alpha_{i+2},\ldots,\alpha_n}, a^{\alpha_1,\ldots,\alpha_i}\}_{yz} \\ &\quad + C^{i+1}_{n-1}\{A^{\alpha_1,\ldots,\alpha_{i+1}}, a^{\alpha_{i+2},\ldots,\alpha_n}\}_{yz}) \end{aligned}$$

and taking the integral with respect to λ we find out that $B^{\alpha_1,\ldots,\alpha_n} = (1/n!)\int a^{\alpha_1,\ldots,\alpha_n}(\lambda)d\lambda$ satisfies the equation for the nth invariant subspace. Finally, formula (13) solves the problem of constructing the solutions of (8) in the case of semisimple algebras, moreover the constructed solutions depend on arbitrary functions, the number of which is equal to the sum of the dimension of the algebra and the dimension of its invariant subspace with the zero graduate index, and gives explicit expressions for each function of the invariant subspace with the finite graduate index.

IV. THE EQUATION IN THE TWO-DIMENSIONAL SPACE AS A RESULT OF THE SUMMATION PERTURBATIVE EXPANSION SERIES

In the present section using the formulas of the previous section we shall sum perturbation series in the case of the algebra $A_1(\mathrm{SL}(2,R))$, and the results will be generalized for the case of arbitrary semisimple algebras.

The result of summation is the following: the summation is equivalent to the solution of a nonlinear integral-differential system of equations with the kernel of Cauchy type defined by operators of the Lax pair. In this case the generators of the first invariant subspace have two indices denoted by + and −, and the system of generators is the same as the solvable part of the infinite-dimensional Kac–Moody algebra $\tilde{A}_2$.

Let us introduce the notation

$$a^{\overbrace{+-\cdots+-}^{2n}} = a^0_n \quad a^{\overbrace{-+\cdots-+}^{2n+1}} = a^+_n,$$

$$a^{\overbrace{-+\cdots+-}^{2n+1}} = a^-_n$$

and

$$a^0 = \tau + \sum_{k=1}^{\infty} \frac{a_k^0}{(2k)!} g^{2k}, \quad u^0 = \frac{\partial a^0}{\partial g},$$
$$a^{\pm} = \sum_{k=0}^{\infty} \frac{a_k^{\pm}}{(2k+1)!} g^{2k+1}, \quad u^{\pm} = \frac{\partial a^{\pm}}{\partial g}. \tag{15}$$

The general recurrent formula in terms of (15) takes the form

$$\frac{\partial u}{\partial g} = [\bar{u}(\lambda) u(\lambda)], \tag{16}$$

where

$$u(\lambda g) = u^+ X^+ + u^0 \lambda + u^- X^-,$$
$$\bar{u}(\lambda, g) = \int d\lambda' \frac{u(\lambda')}{\lambda - \lambda'}.$$

Equation (16) with the initial conditions

$$u(\lambda, 0) = a^1(\lambda)$$

is completely equivalent to the summation of the perturbative expansion series. Thus the constructed solutions are two dimensional by their nature and the interaction constant is an independent variable under this approach. In the notation of (15) we rewrite recurrent relations (14) in the form

$$\frac{\partial^2 a}{\partial g\, \partial \bar{z}} = \lambda \frac{\partial^2 a}{\partial g\, \partial y} + \left[H_y \frac{\partial a}{\partial g} \right] + \left[\frac{\partial a}{\partial g} \int a_y\, d\lambda \right],$$
$$\frac{\partial^2 a}{\partial g\, \partial \bar{y}} = -\lambda \frac{\partial^2 a}{\partial g\, \partial z} - \left[H_z \frac{\partial a}{\partial g} \right] - \left[\frac{\partial a}{\partial g} \int a_z\, d\lambda \right]. \tag{17}$$

Relations (17) are obtained from (14) in the case of the algebra A_1, however, it can take place in the case of an arbitrary algebra. In fact, from (17) it follows that

$$\frac{\partial}{\partial g} (\Box A + [H_y + A_y,\ H_z + A_z]) = 0,$$

i.e., if at $g = 0$ Eq. (8) is fulfilled then it will be fulfilled at arbitrary values of the interaction constant. Thus the general scheme of summation of the perturbative expansion series in the case of the arbitrary semisimple algebra consists of solving Eq. (16) under the initial conditions

$$a(g, \lambda)|_{g=0} = a(\lambda),$$

where $a(\lambda)$ is the solution given by formulas (11) of Eq. (8) for the solvable part of the corresponding semisimple algebra, the nilpotent part of which is the same as the space of positive roots and the zero subspace is pulled over the elements of the Cartan subalgebra.

V. CONCLUSION

The main result of the present work consists of formula (13), which gives an explicit expression for the coefficient functions of the solutions of (8) in the case of an arbitrary solvable algebra, and of the assertion that the summation of the expansion of the perturbative series is equivalent to the solution of the integral-differential equation (16). The work imposes a number of questions. The first question is what class of solutions is described by the proposed scheme, because it can be increased in the sense of reserving arbitrary functions to the general solution.

From a physical point of view this is the question about choosing the interacting constant and about using this choice in physical supplements. Moreover the quadratures on the single parameter suggest a correspondence with string theories.

ACKNOWLEDGMENTS

The authors thank B. A. Arbuzov, O. A. Khrustalev, A. A. Kirillov, Yu. I. Manin, M. A. Mestvirishvili, M. V. Saveliev, A. B. Shabat, and V. I. Man'ko for helpful discussions.

APPENDIX A: A GENERAL SOLUTION OF DALAMBER'S EQUATION

In this appendix a general solution of Dalamber's equation will be obtained. The equation under consideration is the following:

$$\Box u = \frac{\partial^2 u}{\partial y\, \partial \bar{y}} + \frac{\partial^2 u}{\partial z\, \partial \bar{z}} = 0. \tag{A1}$$

For this equation let us introduce two Bäcklund transformations connecting the solutions f and $\tilde{f}$, g and $\tilde{g}$, respectively,

$$f_z = \tilde{f}_{\bar{y}}, \quad g_{\bar{z}} = \tilde{g}_y, \quad f_y = -\tilde{f}_{\bar{z}}, \quad g_y = -\tilde{g}_{\bar{z}}.$$

If we assume that $\tilde{f} = \lambda f$, $\tilde{g} = \lambda g$, where λ is an arbitrary parameter, then we can rewrite these systems in the form

$$f_z = \lambda f_{\bar{y}}, \quad g_{\bar{z}} = \lambda g_{\bar{y}}, \quad f_y = -\lambda f_{\bar{z}}, \quad g_y = -\lambda g_z.$$

Solving these systems we find out that their solutions are

$$f = f(\lambda z + \bar{y}, \lambda y - \bar{z}, \lambda), \quad g = g(\lambda \bar{z} + \bar{y}, \lambda y - z, \lambda),$$

respectively, where f and g are arbitrary functions. Due to the fact that the equation under consideration is a linear one and λ is an arbitrary constant, the sum, and, consequently the integral with respect to λ of the solutions is the solution of Eq. (8), too. Finally, the general solution of the four-dimensional Dalamber equation

$$u = \int d\lambda\, [f(\lambda z + \bar{y}, \lambda y - \bar{z}, \lambda) + g(\lambda \bar{z} + \bar{y}, \lambda y - z, \lambda)]$$

depends on two functions of three independent variables.

APPENDIX B: PROOF OF THE MAIN REDUCTION FORMULAS (14) BY INDUCTION

The main reduction formulas (14) will be proved by induction. From (10) for $n = 1$ we have

$$a_z^1 = \lambda a_y^1 + \rho_g^1 a^1.$$

For $n = 2$

$$a_{\bar{z}}^n = [a^1 a^2]_{\bar{z}}^1 = (\lambda a_y^1(\lambda) + \rho_y^1 a^1(\lambda)) \int \frac{a^2(\lambda') d\lambda'}{\lambda - \lambda'} + a^1(\lambda) \int \frac{(\lambda' - \lambda + \lambda) a_y^2(\lambda') + \rho_y^2 a^2(\lambda')}{\lambda - \lambda'} d\lambda'$$
$$- (\lambda a_y^2(\lambda) + \rho_y^2 a^2(\lambda)) \int \frac{a^1(\lambda') d\lambda'}{\lambda - \lambda'} - a^2(\lambda) \int \frac{(\lambda' - \lambda + \lambda) a_y^1(\lambda') + \rho_y^1 a^1(\lambda')}{\lambda - \lambda'} d\lambda'$$
$$= \lambda [a^1 a^2]_y^1 + (\rho_y^1 + \rho_y^2)[a^1, a^2]^1 - a^1 A_y^2 + a^2 A_y^1 = \lambda a_y^{12} + (\rho_y^1 + \rho_y^2) a^{12} - a^1 A_y^2 + a^2 A_y^1,$$

where $A^\alpha = \int a^\alpha(\lambda)d\lambda$. It is evident that the expression obtained corresponds to the first formula in (14) when $n = 2$. We will consider that it takes place for all orders less than n. Let us consider the expression $[a^{1,\dots,k},a^{k+1,\dots,n}]^1_{\bar z}$. Acting as in the case $n = 2$ and using the supposition of induction we obtain for this expression the following presentation:

$$[a^{1,\dots,k},a^{k+1,\dots,n}]^1_{\bar z} = \lambda\,[a^{1,\dots,k},a^{k+1,\dots,n}]^1_y + \sum_{i=0}^{n} \rho^i_y [a^{1,\dots,k},a^{k+1,\dots,n}]$$
$$- a^{1,\dots,k}A^{k+1,\dots,n}_y + a^{k+1,\dots,n}A^{1,\dots,k}_y + \left[\sum_{\alpha=1}^{k-1} C^\alpha_{k-1} a^{\alpha+1,\dots,k}A^{1,\dots,\alpha}_y - C^{\alpha-1}_{k-1} a^{1,\dots,\alpha}A^{\alpha+1,\dots,k}_y, a^{k+1,\dots,n}\right]^1$$
$$+ \left[a^{1,\dots,k}, \sum_{\alpha=1}^{n-k-1} C^\alpha_{n-k-1} a^{k+1+\alpha,\dots,n}A^{k+1,\dots,k+\alpha}_y - C^{\alpha-1}_{n-k-1} a^{k+1,\dots,k+\alpha}A^{k+\alpha+1,\dots,n}_y\right]^1. \tag{B1}$$

From (13), $\Sigma^{n-1}_{k=1} C^{k-1}_{n-2}[a^{1,\dots,k},a^{k+1,\dots,n}]^1 = a^{1,\dots,n}$. Multiplying both parts of (B1) by C^{k-1}_{n-2} and taking the sum from 1 to $n-1$ we obtain for the last four items of the right part of (B1) the following expressions:

$$\sum_{k=2}^{n-1}\sum_{\alpha=1}^{k-1} C^{k-1}_{n-2}C^{\alpha-1}_{k-1}[a^{1,\dots,\alpha},a^{\alpha+1,\dots,n}]^1 A^{\alpha+1,\dots,k}_y, \tag{B2}$$

$$\sum_{k=1}^{n-2}\sum_{\alpha=1}^{n-k-1} C^{k-1}_{n-2}C^\alpha_{n-k-1}[a^{1,\dots,k},a^{k+1+\alpha,\dots,n}]^1 A^{k+1,\dots,k+\alpha}_y, \tag{B3}$$

$$\left(\sum_{k=2}^{n-1}\sum_{\alpha=1}^{k-1} C^{k-1}_{n-2}C^\alpha_{k-1}[a^{\alpha+1,\dots,k},a^{k+1,\dots,n}]^1 A^{1,\dots,k}_y\right) + \sum_{k=1}^{n-1} C^{k-1}_{n-2}a^{k+1,\dots,n}A_y, \tag{B4}$$

$$\left(\sum_{k=1}^{n-2}\sum_{\alpha=1}^{n-k-1} C^{k-1}_{n-2}C^{\alpha-1}_{n-k-1}[a^{1,\dots,k},a^{k+1,\dots,k+\alpha}]^1 A^{k+\alpha+1,\dots,n}_y\right) + \sum_{k=1}^{n-1} C^{k-1}_{n-2}a^{1,\dots,k}A^{k+1,\dots,n}_y. \tag{B5}$$

Substituting $m = k + \alpha$ in (B3) and using the evident property of binomial coefficients,

$$C^{k-1}_{n-2}C^{m-k}_{n-k-1} = C^{m-1}_{n-2}C^{k-1}_{m-1},$$

one gets convinced that (B3) takes the form

$$\sum_{m=2}^{n-1}\sum_{k=1}^{m-1} C^{m-1}_{n-2}C^{k-1}_{m-1}[a^{1,\dots,k},a^{m+1,\dots,n}]^1 A^{k+1,\dots,m}_y,$$

i.e., the items (B2) and (B3) are being reduced. Let us consider now expression (B4). Carrying out the substitution $k = m + \alpha + 1$ in the expression in brackets we rewrite it in the following form:

$$\sum_{\alpha=1}^{n-2}\sum_{n=0}^{n-\alpha-2} C^{m+\alpha}_{n-2}C^\alpha_{m+\alpha} \times [a^{\alpha+1,\alpha+1+m},a^{\alpha+1+m+1,\dots,k}]^1 A^{1,\dots,\alpha}_y. \tag{B6}$$

Using the property of binomial coefficients,

$$C^{m+\alpha}_{n-2}C^\alpha_{m+\alpha} = C^\alpha_{n-2}C^m_{n-\alpha-2},$$

and definition (13),

$$\sum_{m=0}^{n-m-2} C^m_{n-\alpha-2} \times [a^{\alpha+1,\dots,\alpha+1+m},a^{\alpha+1,m+1,\dots,n}]^1 = a^{\alpha+1,\dots,n},$$

one transforms (B6) to the form $\Sigma^{n-2}_{\alpha=1} C^\alpha_{n-2}a^{\alpha+1,\dots,n}A^{1,\dots,\alpha}_y$. Thus we arrive at the final expression for (B4):

$$\sum_{\alpha=1}^{n-2} C^\alpha_{n-2}a^{\alpha+1,\dots,n}A^{1,\dots,\alpha}_y + \sum_{k=1}^{n-1} C^{k-1}_{n-2}a^{k+1,\dots,n}A^{1,\dots,k}_y$$
$$= \sum_{k=1}^{n-2}(C^k_{n-2} + C^{k-1}_{n-2})a^{k+1,\dots,n}A^{1,\dots,k}_y + C^{n-2}_{n-2}a^n A^{1,\dots,n-1}_y$$
$$= \sum_{k=1}^{n-1} C^{k-1}_{n-1}a^{1,\dots,k}A^{k+1,\dots,n}_y.$$

Similarly one can get the following expression for (B5):

$$\sum_{k=1}^{n-1} C^{k-1}_{n-1}a^{1,\dots,k}A^{k+1,\dots,n}_y.$$

Consequently we have at arbitrary n,

$$a^{1,\dots,n}_{\bar z} = \lambda a^{1,\dots,n}_y + \sum_{i=1}^{n}\rho^i_y a^{1,\dots,n} + \sum_{k=1}^{n-1}(C^k_{n-1}a^{k+1,\dots,n}A^{1,\dots,k}_y - C^{k-1}_{n-1}a^{1,\dots,k}A^{k+1,\dots,n}_y),$$

which completes the proof.

[1]M. K. Prasad, Physica D **1**, 167 (1980).
[2]A. N. Leznov, in *Proceedings of the International Workshop on Nonlinear and Turbulent Processes* (Gordon and Breach, New York, 1984).
[3]A. N. Leznov, V. G. Smirnov, and A. B. Shabat, Theor. Math. Phys. **51**, 10 (1982).
[4]V. G. Kac, Aiz. Akad. Nauk USSR, ser. Math. **32**, 1323 (1968).

On a new class of completely integrable nonlinear wave equations. II. Multi-Hamiltonian structure

Y. Nutku
TUBITAK Research Institute for Basic Sciences, Gebze, Kocaeli, Turkey and Bosphorus University, Bebek, Istanbul, Turkey

(Received 18 December 1985; accepted for publication 8 April 1987)

The multi-Hamiltonian structure of a class of nonlinear wave equations governing the propagation of finite amplitude waves is discussed. Infinitely many conservation laws had earlier been obtained for these equations. Starting from a (primary) Hamiltonian formulation of these equations the necessary and sufficient conditions for the existence of bi-Hamiltonian structure are obtained and it is shown that the second Hamiltonian operator can be constructed solely through a knowledge of the first Hamiltonian function. The recursion operator which first appears at the level of bi-Hamiltonian structure gives rise to an infinite sequence of conserved Hamiltonians. It is found that in general there exist two different infinite sequences of conserved quantities for these equations. The recursion relation defining higher Hamiltonian structures enables one to obtain the necessary and sufficient conditions for the existence of the $(k+1)$st Hamiltonian operator which depends on the kth Hamiltonian function. The infinite sequence of conserved Hamiltonians are common to all the higher Hamiltonian structures. The equations of gas dynamics are discussed as an illustration of this formalism and it is shown that in general they admit tri-Hamiltonian structure with two distinct infinite sets of conserved quantities. The isothermal case of $\gamma = 1$ is an exceptional one that requires separate treatment. This corresponds to a specialization of the equations governing the expansion of plasma into vacuum which will be shown to be equivalent to Poisson's equation in nonlinear acoustics.

I. INTRODUCTION

The demonstration of a bi-Hamiltonian structure for a system of partial differential equations is a direct and elegant method of proving its complete integrability.[1–5] That is, if a set of partial differential equations can be formulated as a Hamiltonian system in two distinct but compatible ways, then, by a theorem of Magri,[2] they give rise to an infinite sequence of conserved Hamiltonians which are in involution with respect to either one of these two symplectic structures. It is well known that nonlinear evolution equations which can be solved by the inverse scattering method[6,7] can be cast into bi-Hamiltonian form.[2] In a recent paper,[8] hereafter to be denoted as I, we have pointed out a new class of nonlinear wave equations that do not admit a nontrivial Lax pair but nevertheless give rise to infinitely many conservation laws. Equations governing the propagation of finite amplitude waves such as the Euler and Poisson equations in nonlinear acoustics, shallow water waves, the simplest nonlinear Born–Infeld electrodynamics, the Morse–Ingard string, and Nambu's relativistic string belong to this class. We had earlier presented a general framework for these equations and gave an algorithm for constructing infinitely many conservation laws. These conserved quantities are Casimir functions[9] and in this paper we shall show that they arise as a consequence of the bi-Hamiltonian structure of some members of this class of nonlinear wave equations.

The discussion of bi-Hamiltonian structure suggests a generalization of the recursion relation which enables us to obtain the necessary and sufficient conditions for the existence of higher Hamiltonian structures and construct the appropriate recursion operators. In the case of the equations governing the propagation of long waves in shallow water, or gas dynamics with $\gamma = 2$, the second Hamiltonian operator was obtained by Cavalcante and McKean[10] and Kupershmidt[11] has shown that these equations admit tri-Hamiltonian structure. The infinite sequence of conserved Hamiltonian that this structure gives rise to was found much earlier by Benney.[12] It appears that Kupershmidt's treatment of this problem and its generalization to include the effects of dispersion[13] are the only discussions of multi-Hamiltonian structure in the literature. We shall show that it is not possible to go beyond tri-Hamiltonian structure for shallow water waves. We shall discuss the equations of gas dynamics in $1+1$ dimensions as an illustration of our formalism for multi-Hamiltonian structure. We shall find that in the generic case these equations admit tri-Hamiltonian structure with two infinite sequences of conserved quantities. For $\gamma = 2$ these two sequences collapse into one yielding Benney's conserved quantities. Physically the most interesting case occurs for an isothermal gas with $\gamma = 1$. We shall show that the equations are now equivalent to Poisson's equation[14] in nonlinear acoustics and they are of further interest in plasma physics.[15,16] Mathematically this is an exceptional case which we shall treat separately.

II. SYMPLECTIC STRUCTURE

A. Primary Hamiltonian structure

In I we had considered quasilinear second-order partial differential equations

$$F(\phi_t, \phi_x, \phi_{tt}, \phi_{tx}, \phi_{xx}) = 0, \tag{2.1}$$

which can be written in the form of a continuity equation.

These equations can alternatively be formulated in terms of a pair of exact one-forms

$$\alpha = u\,dx + U(u,v)dt = d\phi, \quad \omega = v\,dx + V(u,v)dt = d\psi, \tag{2.2}$$

and the implication that they are closed gives rise to the first-order equations

$$u_t = U_u u_x + U_v v_x \equiv K^u, \quad v_t = V_u u_x + V_v v_x \equiv K^v, \tag{2.3}$$

which consist of a restatement of Eq. (2.1) and further result in a companion equation for ψ. There is, in general, an ambiguity in the choice of ψ due to the fact that different sets of coupled first-order partial differential equations for ϕ,ψ, which are obtained as compatibility conditions of Eqs. (2.2), may yield Eq. (2.1) as *one* of their integrability conditions. This ambiguity is resolved precisely by requiring that the system must admit Hamiltonian structure, for which

$$U_u = V_v \tag{2.4}$$

is the necessary condition.

We recall that in the discussion of the Hamiltonian structure of this class of equations we start with Eqs. (2.3) as the basic equations. The phase space consists of the set (u,v) of infinitely differentiable functions and the Euler operator

$$E = \begin{pmatrix} E_u \\ E_v \end{pmatrix}, \quad E_u = \frac{\delta}{\delta u}, \quad E_v = \frac{\delta}{\delta v}, \tag{2.5}$$

where δ denotes the variational derivative, will stand for the gradient in this space. For two smooth functions A, B of these variables the Poisson bracket is defined by

$$[A,B] = \int E(A) J E(B) dx, \tag{2.6}$$

where J is the Hamiltonian operator. Hamilton's equations are given by

$$\frac{\partial}{\partial t}\begin{pmatrix} u \\ v \end{pmatrix} = \left[\begin{pmatrix} u \\ v \end{pmatrix}, H\right] = JE(H), \tag{2.7}$$

with H standing for the Hamiltonian function. We had found that for the familiar Hamiltonian operator[3,8,9]

$$J_1 = -\begin{pmatrix} 0 & D \\ D & 0 \end{pmatrix}, \quad D = \frac{\partial}{\partial x}, \tag{2.8}$$

the necessary and sufficient conditions for the existence of a Hamiltonian function H_1 are given by the requirement that

$$-dH_1 = U\,dv + V\,du \tag{2.9}$$

must be an exact one-form. Equation (2.4) is obtained by taking the exterior derivative of Eq. (2.9). This is the first Hamiltonian structure. It is evident that the proper choice of ψ is the most crucial element in the discussion of the Hamiltonian structure of the nonlinear wave equations that belong to the class considered in I.

B. Bi-Hamiltonian structure

We shall now consider the problem of constructing the second Hamiltonian structure of Eqs. (2.3). There exist excellent accounts of bi-Hamiltonian structure in the literature.[17–19] Since our approach is tailored to a specific class of equations, various aspects of this construction can be understood intuitively and in the following we shall make extensive use of these simplifications. For this purpose we start by noting that as a consequence of Eqs. (2.3) and (2.4),

$$H_0 = uv \tag{2.10}$$

is always a conserved quantity. That is, Eq. (2.4) is the integrability condition of $H_{0t} + K_{0x} = 0$, where K_0 is a function of u, v only. Thus we may inquire into the existence of a second Hamiltonian operator J_2, such that

$$J_2 E(H_0) = J_1 E(H_1) \tag{2.11a}$$

and Eqs. (2.3) will become Hamilton's equations with respect to the Poisson brackets defined in terms of both J_1 and J_2. In order to determine J_2 it is necessary to require that every linear combination of J_1 and J_2 with constant coefficients is also a Hamiltonian operator. The existence of such a bi-Hamiltonian structure gives rise to a new conserved Hamiltonian H_2 satisfying

$$J_2 E(H_1) = J_1 E(H_2) \tag{2.11b}$$

and the recursion relation

$$J_2 E(H_{k-1}) = J_1 E(H_k) \tag{2.11c}$$

enables us to obtain an infinite sequence of conserved quantities H_k for every integer k.

A Hamiltonian operator must be skew adjoint. That is,

$$\int_a^b GJF\,dx = -\int_a^b FJG\,dx \tag{2.12}$$

and this is accomplished through integration by parts and suitable boundary conditions at the limits of integration. For the class of nonlinear wave equations discussed in I, J is a 2×2 matrix and F, G are each two-component objects. Direct calculation shows that the most general ansatz for J which is consistent with Eq. (2.12) is given by

$$J_2 = \begin{pmatrix} mD + Dm & (p-q)D + D(p+q) \\ (p+q)D + D(p-q) & nD + Dn \end{pmatrix}, \tag{2.13}$$

where m, n, p, and q are functions of u, v. We note that the diagonal entries of J_2 consist of first-order manifestly antisymmetric operators while the off-diagonal terms contain a first- and a zeroth-order operator involving p, q, respectively, where it is the change in the sign of q between the off-diagonal elements of J_2 that assures antisymmetry in the latter case. The task of checking whether or not Eq. (2.12) is satisfied for a given set of boundary conditions belongs to the end, after the functions entering into this ansatz are completely known for each particular example. These functions, $\{m,n,p,q\}$, must be determined by requiring the satisfaction of Eqs. (2.11a) and the Jacobi identity. This is a straightforward but lengthy calculation which we shall now outline.

Equations (2.11a) result in four equations for the first derivatives of U,V and from the equality of the mixed second derivatives of these functions and Eq. (2.4) we find

$$m_u = p_v - q_v, \tag{2.14a}$$

$$n_v = p_u + q_u, \tag{2.14b}$$

which enables us to write

$$2p + un_v + vm_u = U_u \tag{2.14c}$$

in place of two of these equations. The content of the remaining two equations is reduced to

$$U_v(un)_u + V_u(vm)_v = U_v V_u \tag{2.14d}$$

by virtue of Eqs. (2.11b), which consists of relations for the second derivatives of H_2. That is, from Eqs. (2.11b) we find

$$nU_v = mV_u, \tag{2.14e}$$

$$U_v n_u = V_u(p_v + q_v), \tag{2.14f}$$

$$V_u m_v = U_v(p_u - q_u), \tag{2.14g}$$

using the equality of the mixed second derivatives of H_2, H_{2u}, and H_{2v}, respectively. Equations (2.14) are an overdetermined set of first-order partial differential equations which must be solved in order to construct the second Hamiltonian operator J_2 when it exists. But before we can proceed with this matter we need to show that they are sufficient to insure the satisfaction of the Jacobi identities.

We shall adopt a formalism due to Olver[19] for calculating the Jacobi identities that appears to be the most direct and economical method devised for this purpose. We consider tangent vectors

$$\Theta = \begin{pmatrix} \theta \\ \zeta \end{pmatrix} \tag{2.15}$$

and the bilinear form

$$I = \int_b^a \mathscr{I}\, dx, \quad \mathscr{I} = \Theta^T J \Theta, \tag{2.16}$$

where products involving θ, ζ, θ_x,... are understood to be completely skew so that they can be manipulated with the same rules governing differential forms. We shall use the symbol $\wedge$ for wedge product in order to make this property manifest. Finally, introducing the operator

$$L = (J\Theta)\cdot E \tag{2.17}$$

the Jacobi identities can be expressed in the form

$$LI = 0, \tag{2.18}$$

where we again assume that suitable boundary conditions assure the vanishing of terms that are to be evaluated at the limits of integration. In our case we find

$$\mathscr{I} = m\theta\wedge\theta_x + p(\theta\wedge\zeta_x + \zeta\wedge\theta_x) + n\zeta\wedge\zeta_x + (q_u u_x + q_v v_x)\theta\wedge\zeta, \tag{2.19a}$$

$$E_u\mathscr{I} = m_u\theta\wedge\theta_x + (p_u - q_u)\theta\wedge\zeta_x + (p_u + q_u)\zeta\wedge\theta_x + n_u\zeta\wedge\zeta_x, \tag{2.19b}$$

and $E_v\mathscr{I}$ is obtained by replacing the subscript u by v everywhere it occurs in Eq. (2.19b). Evaluating Eq. (2.18) we find that there are four linearly independent quantities $\theta\wedge\theta_x\wedge\zeta_x$, $\theta_x\wedge\zeta\wedge\zeta_x$, $\theta\wedge\theta_x\wedge\zeta$, and $\theta\wedge\zeta\wedge\zeta_x$, the coefficients of which must vanish. The latter two each give rise to two equations as they involve terms proportional to u_x and v_x as well, but one of these equations, Eq. (2.20e) below, is repeated. Thus we end up with

$$p[m_u - p_v + q_v] + nm_v - m(p_u - q_u) = 0, \tag{2.20a}$$

$$p[n_v - p_u - q_u] + mn_u - n(p_v + q_v) = 0, \tag{2.20b}$$

$$[n_v - p_u - q_u]m_v + [m_u - p_v + q_v](p_v + q_v) = 0, \tag{2.20c}$$

$$[n_v - p_u - q_u](p_u - q_u) + [m_u - p_v + q_v]n_u = 0, \tag{2.20d}$$

$$m_v n_u - (p_v + q_u)(p_u - q_u) = 0, \tag{2.20e}$$

as the conditions for satisfaction of the Jacobi identities.

Comparing with earlier results, we find that Eqs. (2.20) are completely equivalent to Eqs. (2.14a), (2.14b), (2.14e), (2.14f), and (2.14g). The difference between these two sets of equations lies in the fact that, in contrast to the Jacobi identities (2.20), Eqs. (2.14) contain explicit reference to the Hamiltonian function H_1 of the system for which we are interested in constructing the second Hamiltonian operator. This fact is of crucial importance in finding J_2 for particular examples.

We shall now consider the problem of constructing the second Hamiltonian operator J_2 using Eqs. (2.14). But first of all we need to check the integrability conditions of these equations in order to verify that a solution will exist for some U, V subject to Eq. (2.4). To this end we define a new function f in terms of which Eqs. (2.14) become

$$m = U_v f, \tag{2.21a}$$

$$n = V_u f, \tag{2.21b}$$

$$p = \tfrac{1}{2}[U_u - uV_u f_v - vU_v f_u - (uV_{uv} + vU_{uv})f], \tag{2.21c}$$

$$p_u = V_u f_v + \tfrac{1}{2}(V_{uv} + V_u U_{vv}/U_v)f, \tag{2.21d}$$

$$p_v = U_v f_u + \tfrac{1}{2}(U_{uv} + U_v V_{uu}/V_u)f, \tag{2.21e}$$

$$q_u = \tfrac{1}{2}(V_{uv} - V_u U_{vv}/U_v)f, \tag{2.21f}$$

$$q_v = \tfrac{1}{2}(-U_{uv} + U_v V_{uu}/V_u)f, \tag{2.21g}$$

and a first-order equation for f only

$$uf_u + vf_v + (2 + uV_{uu}/V_u + vU_{vv}/U_v)f = 1. \tag{2.21h}$$

It turns out that Eq. (2.21h) and the integrability condition of Eqs. (2.21f) and (2.21g),

$$[U_v(V_u/U_v)_u f]_v + [V_u(U_v/V_u)_v f]_u = 0, \tag{2.21i}$$

are the crucial equations which determine f and once we find a *particular* solution of this pair of first-order partial differential equations, the required functions m, n, p, q can be obtained by quadratures. An exhaustive check of all the integrability conditions of Eqs. (2.21) is rather lengthy as the following example will illustrate. From Eqs. (2.21c) and (2.21d) we find a second-order equation for f. An equation involving the same second derivatives of f is obtained by differentiating Eq. (2.21h). Combining these equations we end up with an equation for f_{uv} in terms of lower derivatives of f. When this process is repeated with Eqs. (2.21c) and (2.21e) we again find a similar equation for f_{uv} and by comparing it with the earlier result we obtain a first-order equation for f which is simply Eq. (2.21i).

Thus the existence of a second Hamiltonian structure for Eqs. (2.3) with J_2 given by Eqs. (2.13) and the solution of Eqs. (2.21), depends on whether or not Eqs. (2.21h) and (2.21i) admit a solution. This is an algebraic system for the first derivatives of f and its solution yields the Meyer system

$$f_u = -\rho(\beta_u + \gamma_v)vf + \rho\mu\gamma f - \rho\gamma, \tag{2.22a}$$

$$f_v = \rho(\beta_u + \gamma_v)uf - \rho\mu\beta f + \rho\beta, \tag{2.22b}$$

where

$$\beta = V_u(U_v/V_u)_u, \quad \gamma = U_v(V_u/U_v)_v,$$
$$\mu = 2 + uV_{uu}/V_u + vU_{vv}/U_v, \quad \rho = (\beta v - \gamma u)^{-1}. \tag{2.23}$$

The integrability conditions for the Meyer system are well known[20,21]: We consider the one-form

$$\sigma = df - (Af + B), \tag{2.24a}$$

where

$$A = \rho(\beta_u + \gamma_v)(u\,dv - v\,du) - \rho\mu(\beta\,dv - \gamma\,du), \tag{2.24b}$$

$$B = \rho(\beta\,dv - \gamma\,du) \tag{2.24c}$$

depend only on u, v. The necessary and sufficient conditions for the complete integrability of Eq. (2.22) are given by

$$f\,dA + dB - A \wedge B = 0 \tag{2.25}$$

and this implies

$$d\sigma = -\sigma \wedge A, \tag{2.26}$$

where, by the theorem of Frobenius, A is exact,

$$A = d\lambda. \tag{2.27a}$$

Then Eq. (2.25) does not depend on f and λ acts as the integrating factor for the rest. Hence, using Poincaré's lemma, we find

$$B = e^{\lambda} dv \tag{2.27b}$$

and finally

$$f = e^{\lambda} v, \tag{2.28}$$

so that the integration of Eqs. (2.21h) and (2.21i) is reduced to a quadrature.

We conclude that nonlinear wave equations belonging to the class considered in I which further satisfy Eqs. (2.22) have the *recursion operator*

$$\mathscr{D} = J_2 J_1^{-1} \tag{2.29}$$

and admit an infinite sequence of conserved Hamiltonians H_k which are in involution with respect to either one of these two symplectic structures. By a theorem of Olver,[1] the recursion operator satisfies

$$\mathscr{D}_t = [\mathscr{A}, \mathscr{D}] \tag{2.30}$$

with

$$\mathscr{A} = \begin{pmatrix} U_u D + U_{uu} u_x + U_{uv} v_x & U_v D + U_{uv} u_x + U_{vv} v_x \\ V_u D + V_{uu} u_x + V_{uv} v_x & V_v D + V_{uv} u_x + V_{vv} v_x \end{pmatrix}, \tag{2.31}$$

which is the derivative of the vector field K of Eq. (2.3) describing the flow.

Finally we note that the infinite sequence of conserved Hamiltonians that result as a consequence of bi-Hamiltonian structure are anchored in H_0 given by Eq. (2.10). There may be a conserved quantity $\widetilde{H}_0$ which is linearly independent of the set of H_k and once we have the recursion operator we may generate another infinite sequence of conserved quantities $\widetilde{H}_k$ starting with $\widetilde{H}_0$. In particular u, v are conserved by virtue of Eqs. (2.3) and they might be predecessors of H_k or they may serve as anchors of new infinite sequences of conserved quantities which are obtained through the repeated application of the recursion operator.

C. Multi-Hamiltonian structure

In our discussion of the second Hamiltonian operator we have used the requirement that U, V should be derivable from a primary Hamiltonian H_1 through Eq. (2.9) and obtained Eqs. (2.27) for the existence of J_2. Given such a bi-Hamiltonian structure, the third Hamiltonian operator J_3 may be defined through the recursion relation

$$J_3 E(H_k) = J_2 E(H_{k+1}), \tag{2.32a}$$

where, knowing J_2, H_1, H_2, we can determine J_3 even as we found J_2 from a knowledge of J_1, H_0, H_1. It is evident that there is no reason to stop here and we can define infinitely many Hamiltonian structures, provided that they exist of course. The recursion relation is now given by

$$J_l E(H_{k+l-3}) = J_{l-1} E(H_{k+l-2}) \tag{2.32b}$$

for a pair or integers k, l. Repeated use of Eq. (2.32b) enables us to express the general recursion relation in the form

$$J_k E(H_{i+k-3}) = J_l E(H_{i+2k-l-3}), \tag{2.32c}$$

where i, k, l are integers with $k > l$. We can then define recursion operators

$$\mathscr{D}_{kl} = J_k J_l^{-1} = \mathscr{D}_{ki} \mathscr{D}_{il} \tag{2.33}$$

provided that the inverse of these Hamiltonian operators exist. The infinite sequence of conserved Hamiltonians is common to all these Hamiltonian structures. The conserved quantities can be generated through the recursion operator $\mathscr{D}_{kl}$ in steps of $k - l$. In this notation the recursion operator of Eq. (2.29) is $\mathscr{D}_{21}$.

We shall now consider the problem of constructing J_{k+1} for a given H_k. This will be a straightforward generalization of the results of the previous section. In particular we shall start with the ansatz of Eq. (2.13) where the functions m, n, p, q will carry the subscript $k+1$ for J_{k+1} and they will satisfy equations identical to Eqs. (2.14) except that the kth Hamiltonian function will appear in place of the first one. Suppressing cumbersome subscripts, we can summarize our results as follows.

Theorem: Given the kth Hamiltonian function H, the existence of a Hamiltonian operator of order $k+1$ requires that the Meyer system,

$$uf_u + vf_v + (2 + u(H_{uuu}/H_{uu}) + v(H_{vvv}/H_{vv}))f = 1, \tag{2.34a}$$

$$[(H_{vv}/H_{uu})_u H_{uu} f]_u + [(H_{uu}/H_{vv})_v H_{vv} f]_v = 0, \tag{2.34b}$$

admits a solution for f. The necessary and sufficient conditions for the existence of such a solution are given by

$$dA = 0, \tag{2.35a}$$

$$dB - A \wedge B = 0, \tag{2.35b}$$

where

$$A = \frac{[H_{uu}(H_{vv}/H_{uu})_u]_u + [H_{vv}(H_{uu}/H_{vv})_v]_v}{vH_{uu}(H_{vv}/H_{uu})_u - uH_{vv}(H_{uu}/H_{vv})_v} \times (u\,dv - v\,du) - \left(2 + u\frac{H_{uuu}}{H_{uu}} + v\frac{H_{vvv}}{H_{vv}}\right)B, \tag{2.36a}$$

$$B=\frac{H_{uu}(H_{vv}/H_{uu})_u\,dv-H_{vv}(H_{uu}/H_{vv})_v\,du}{vH_{uu}(H_{vv}/H_{uu})_u-uH_{vv}(H_{uu}/H_{vv})_v},\tag{2.36b}$$

and in this case

$$f=\exp\left(\int A\right)\int\left[B\exp\left(-\int A\right)\right]\tag{2.37}$$

can be obtained by quadratures. Then we can construct J_{k+1} according to Eq. (2.13) with

$$m=H_{vv}f,\tag{2.38a}$$

$$n=H_{uu}f,\tag{2.38b}$$

$$p=-\tfrac{1}{2}[H_{uv}-uH_{uu}f_v-vH_{vv}f_u-(uH_{uuv}+vH_{uvv})f],\tag{2.38c}$$

and

$$q=-\frac{1}{2}\int f\left[\left(\frac{H_{uu}}{H_{vv}}\right)_v H_{vv}\,du-\left(\frac{H_{vv}}{H_{uu}}\right)_u H_{uu}\,dv\right]\tag{2.38d}$$

is also reduced to a quadrature by virtue of Eq. (2.34b).

III. EXAMPLES

A. Poisson's equation

Poisson's equation in nonlinear acoustics[14]

$$\phi_{tt}+2\phi_x\phi_{tx}-(1-\phi_x{}^2)\phi_{xx}=0\tag{3.1}$$

cannot be given a variational formulation in terms of the velocity potential ϕ alone. Nevertheless, Poisson's equation admits Hamiltonian structure which, due to a poor choice of ψ, we had missed in our earlier discussion of this equation. Thus we consider the first-order system

$$\phi_t+\tfrac{1}{2}\phi_x{}^2+\ln\psi_x=0,\quad \psi_t+\phi_x\psi_x=0,\tag{3.2}$$

the integrability conditions of which yield Eq. (3.1) and

$$\psi_x{}^2\psi_{tt}-2\psi_x\psi_t\psi_{tx}+(\psi_t{}^2-\psi_x{}^2)\psi_{xx}=0,\tag{3.3}$$

which is the proper companion to Poisson's equation. From Eqs. (3.2), (I.2.9), and (I.2.10) we find

$$\phi_x=u,\quad \phi_t=U=-\tfrac{1}{2}u^2-\ln v,$$
$$\psi_x=v,\quad \psi_t=V=-uv,\tag{3.4}$$

whereby the necessary condition (2.4) for the existence of Hamiltonian structure is fulfilled. Then the exact one-forms (2.2) are given by

$$\alpha=u\,dx-(\tfrac{1}{2}u^2+\ln v)dt,\quad \omega=v\,dx-uv\,dt,\tag{3.5}$$

and their integrability conditions yield

$$u_t+uu_x+v^{-1}v_x=0,\quad v_t+vu_x+uv_x=0,\tag{3.6}$$

which are basic equations. These equations govern the expansion of plasma into vacuum when the electrons are assumed to be in isothermal equilibrium, the electron density given by the Boltzmann relation and further assuming quasineutrality and ignoring collisions.[16] The characteristics of Eqs. (3.6) satisfy

$$x'=(u+\epsilon)t',\tag{3.7}$$

with $\epsilon=\pm1$, and $D^\epsilon=\partial_t+(u+\epsilon)\partial_x$ is the derivative along these directions. The Riemann invariants are

$$R^\epsilon=u+\epsilon\ln v\tag{3.8}$$

and the basic equations (3.6) are cast into Riemann's canonical form $D^\epsilon R^\epsilon=0$.

The first Hamiltonian of Poisson's equation,

$$H_1=\tfrac{1}{2}u^2v+v(\ln v-1),\tag{3.9}$$

follows from the integration of Eq. (2.9). We find that Eqs. (2.35) are satisfied for this H_1 so that Poisson's equation admits bi-Hamiltonian structure. Then the solution of Eqs. (2.38) yields the second Hamiltonian operator

$$J_2=\begin{pmatrix}2v^{-1}D-v^{-2}v_x & u_x\\ -u_x & 2vD+v_x\end{pmatrix}\tag{3.10}$$

and through the use of the recursion operator (2.29), which is constructed from Eqs. (2.8) and (3.10), we obtain the infinite sequence of conserved Hamiltonians. The first few of these Hamiltonians are

$$H_2=\tfrac{1}{6}u^3v+uv(\ln v+1),$$
$$H_3=\tfrac{1}{24}u^4v+\tfrac{1}{2}u^2v(\ln v+3)+\tfrac{1}{2}v\ln v(\ln v+4)-2v.\tag{3.11}$$

If we extend the recursion relation (2.11c) to $k=0$, we find that

$$H_{-1}=v\tag{3.12}$$

emerges as a member of this infinite sequence but further extension to negative integer values of k is vacuous. Since u is not included in this set we can take it as the starting point of a new sequence of conserved quantities $\widetilde{H}_k$. Repeated application of the recursion operator yields

$$\widetilde{H}_0=u,\quad \widetilde{H}_1=\tfrac{1}{2}u^2-\ln v-1,$$
$$\widetilde{H}_2=\tfrac{1}{6}u^3-u(\ln v+1),\tag{3.13}$$

which are the first three elements of an infinite sequence.

The conserved Hamiltonians (3.11) and (3.13) are solutions of Eqs. (I.2.27)

$$G_u-uF_u-vF_v=0,\quad G_v-v^{-1}F_u-uF_v=0,\tag{3.14}$$

and the second-order decoupled equation

$$F_{uu}-v^2F_{vv}=0\tag{3.15}$$

is useful for finding the general expression for both of these infinite sequences of conserved quantities. First we shall rewrite Eq. (3.15) as

$$4F_{\xi\eta}+F_\xi-F_\eta=0\tag{3.16}$$

using the Riemann invariants (3.8) as new coordinates with $\xi=R^+,\eta=R^-$. In this form it is manifest that if $\widetilde{F}(\xi,\eta)$ is a solution, then so is

$$F(\xi,\eta)=e^{(1/2)(\xi-\eta)}\widetilde{F}(\eta,\xi)\tag{3.17}$$

and this fact is responsible for the existence of the two sets (3.11) and (3.13). The solution of the equation for the conserved quantities gives

$$\widetilde{F}_{k-1}=\sum_{l=0}^{[k/2]}\frac{(-1)^l}{(k-2l)!}u^{k-2l}P_l(\ln v),\tag{3.18}$$

where $[k/2]$ denotes the largest integer less than $k/2$ and the P_l are polynomials with $P_0=1$,

$$P_{l+1}(y)=\sum_{j=-1}^{l-1}\frac{d^j}{dy^j}P_l(y)\tag{3.19}$$

so that

$$P_1 = y+1, \quad P_2 = \tfrac{1}{2}y^2 + 2y + 2,$$
$$P_3 = \tfrac{1}{6}y^3 + \tfrac{3}{2}y^2 + 5y + 5, \tag{3.20}$$

and so on. From Eq. (3.17) it follows that

$$F_{k-1} = v \sum_{l=0}^{[k/2]} \frac{(-1)^l}{(k-2l)!} u^{k-2} P_l(-\ln v) \tag{3.21}$$

is a new set of conserved quantities. Comparison of Eqs. (3.18) and (3.21) with Eqs. (3.13) and (3.11) shows that

$$H_k = F_k \mod F_j, \ j<k. \tag{3.22}$$

Poisson's equation admits tri-Hamiltonian structure because Eqs. (2.34) admit a solution for H_2 given by Eq. (3.11). The third Hamiltonian operator is

$$J_3 = \begin{pmatrix} (u/v)D + u_x/2v - uv_x/2v^2 & (\ln v + \tfrac{3}{2})D + \tfrac{1}{2}uu_x + v_x/2v \\ (\ln v + \tfrac{3}{2})D - \tfrac{1}{2}uu_x + v_x/2v & uvD + \tfrac{1}{2}vu_x + \tfrac{1}{2}uv_x \end{pmatrix} \tag{3.23}$$

and this is the last one because Eqs. (2.35) are not satisfied for H_3 given in Eqs. (3.11). We note that J_3 is the second Hamiltonian operator for

$$u_t + (\tfrac{1}{2}u^2 + \ln v + 2)u_x + (u/v)v_x = 0,$$
$$v_t + uvu_x + (\tfrac{1}{2}u^2 + \ln v + 2)v_x = 0, \tag{3.24}$$

which, after Eqs. (3.6) themselves, is the next set of equations in the Poisson hierarchy. Running the arguments of Sec. II A backwards, we find that the second-order quasilinear equation

$$\phi_x{}^2\phi_{tt} - 2(\phi_t - \tfrac{1}{3}\phi_x{}^3)\phi_x\phi_{xt} + [(\phi_t - \tfrac{1}{3}\phi_x{}^3)^2 - \phi_x{}^4]\phi_{xx} = 0 \tag{3.25}$$

is the second completely integrable equation following Poisson's equation (3.1).

B. Gas dynamics in 1+1 dimensions

We shall now consider the multi-Hamiltonian structure that the equations of gas dynamics give rise to. In this connection we note that Poisson's equation corresponds to the isothermal case $\gamma = 1$ which is excluded in what follows.

In order to discuss the equations of gas dynamics in the framework of I, we need to use

$$v = \left(\frac{\partial p}{\partial \rho}\right)^{1/\gamma} \tag{3.26}$$

rather than the velocity of sound as one of the basic variables, cf. Appendix I of Ref. 22. Then we have the exact forms

$$\alpha = u\,dx - [\tfrac{1}{2}u^2 + (\gamma-1)^{-1}v^{\gamma-1}]dt,$$
$$\omega = v\,dt - uv\,dt \tag{3.27}$$

and the equations of gas dynamics assume the form

$$u_t + uu_x + v^{\gamma-2}v_x = 0, \quad v_t + vu_x + uv_x = 0, \tag{3.28}$$

which is of the form of Eqs. (2.3).

The first Hamiltonian function for the equations of gas dynamics is given by

$$H_1 = \tfrac{1}{2}u^2 v + \gamma^{-1}(\gamma-1)^{-1}v^\gamma \tag{3.29}$$

and it is the first in an infinite sequence of conserved quantities. Equations (2.35) are satisfied for this Hamiltonian function and therefore we can construct the second Hamiltonian operator J_2 which yields

$$m_2 = \gamma^{-1}v^{\gamma-2}, \quad n_2 = \gamma^{-1}v,$$
$$p_2 = \tfrac{1}{2}\gamma^{-1}(\gamma-1)u, \quad q_2 = \tfrac{1}{2}\gamma^{-1}(3-\gamma)u. \tag{3.30}$$

Using the recursion operator (2.29) we can generate the rest of the infinite sequence of conserved Hamiltonians[23]

$$H_2 = \tfrac{1}{6}u^3 v + \gamma^{-1}(\gamma-1)^{-1}uv^\gamma,$$
$$H_3 = \tfrac{1}{24}u^4 v + \tfrac{1}{2}\gamma^{-1}(\gamma-1)^{-1}u^2v^\gamma + \tfrac{1}{2}\gamma^{-1}(\gamma-1)^{-2}(2\gamma-1)^{-1}v^{2\gamma-1}, \tag{3.31}$$

and so on. However, as in the case of Poisson's equation we find that

$$H_{-1} = \gamma v \tag{3.32}$$

and u is not included in the hierarchy of conserved Hamiltonians. We can therefore use it to generate a new infinite sequence of conserved quantities, the first few of which are

$$\tilde{H}_{-1} = \gamma u, \quad \tilde{H}_0 = \tfrac{1}{2}(\gamma-2)u^2 + (\gamma-1)^{-1}v^{\gamma-1},$$
$$\tilde{H}_1 = (2\gamma-3)(3\gamma-4)\gamma^{-2}[\tfrac{1}{24}(\gamma-2)u^4 + \tfrac{1}{2}(\gamma-1)^{-1}u^2v^{\gamma-1} + \tfrac{1}{2}(\gamma-1)^{-2}(2\gamma-3)^{-1}v^{2\gamma-2}]. \tag{3.33}$$

We note that for $\gamma = 2$, the case of shallow water waves, the two hierarchies of conserved quantities degenerate into one:

$$H_k = \tilde{H}_{k+1}, \quad k \geqslant 0, \quad \gamma = 2. \tag{3.34}$$

The equations of gas dynamics admit tri-Hamiltonian structure. The functions that enter into J_3 are given by

$$m_3 = (1+\gamma)^{-1}uv^{\gamma-2}, \quad n_3 = (1+\gamma)^{-1}uv,$$
$$p_3 = \tfrac{1}{4}(1+\gamma)^{-1}(\gamma-1)u^2 + (\gamma^2-1)^{-1}v^{\gamma-1}, \tag{3.35}$$
$$q_3 = \tfrac{1}{4}(1+\gamma)^{-1}(3-\gamma)u^2,$$

and, apart from the case of $\gamma = 3$, Eqs. (2.35) are not satisfied for H_k, $k \geqslant 3$, so that there exist no further Hamiltonian structures. As we have noted earlier, for the special case of shallow water waves ($\gamma = 2$) the results of Eqs. (3.30) and (3.34) were first obtained in Refs. 10 and 11, respectively.

The case $\gamma = 3$ is noteworthy in that $q_2 = q_3 = 0$ and in fact there exist infinitely many Hamiltonian operators J_k with

$$q_k = 0,$$

$$p_k = \frac{k(2k-1)!}{2^k(k+1)!} \sum_{l=0}^{k-1} \frac{[1+(-1)^{k-1-l}]}{l!(k-1-l)!} u^l v^{k-1-l},$$

$$m_k = n_k = \frac{k(2k-1)!}{2^k(k+1)!} \sum_{l=0}^{k-1} \times \frac{[1-(-1)^{k-1-l}]}{l!(k-1-l)!} u^l v^{k-1-l}, \quad (3.36)$$

but $\gamma = 3$ is not an interesting example of Eqs. (3.28) because they decouple in the variables $u \pm v$. We note that if $q_k = 0$, then Eq. (2.34b) is an identity and Eq. (2.24a) will always admit a solution. Hence there will be infinitely many Hamiltonian structures.

IV. CONCLUSION

The principal result of this paper is the theorem at the end of Sec. II which contains the necessary and sufficient conditions for the existence of a Hamiltonian operator J_{k+1} given a Hamiltonian function H_k. We have also reduced to quadratures the problem of determining the functions that enter into the definition of J_{k+1} once these conditions are satisfied. This result settles the question as to when a nonlinear evolution equation that belongs to the general class discussed in I admits multi-Hamiltonian structure and provides a simple algorithm for constructing the Hamiltonian operator as well. We have demonstrated its utility in the applications to Poisson's equation and the equations of gas dynamics.

ACKNOWLEDGMENTS

I thank M. Gürses, F. Neyzi, P. J. Olver, and J. Verosky for many interesting discussions and correspondence. I would also like to thank Professor W. Miller Jr., and Professor P. J. Olver for an invitation to visit the University of Minnesota and I am grateful to them for their hospitality in Minneapolis.

This work was in part supported by NATO Research Grant No. RG 86/0055.

[1]P. J. Olver, J. Math. Phys. **18**, 1212 (1977).
[2]F. Magri, J. Math. Phys. **19**, 1156 (1978).
[3]I. M. Gel'fand and I. Ya. Dorfman, Funct. Anal. Appl. **15**, 23 (1981).
[4]F. Fuchsteiner and A. S. Fokas, Physica (Utrecht) D **4**, 47 (1981).
[5]P. J. Olver, Contemp. Math. **28**, 231 (1984).
[6]C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[7]V. E. Zakharov and A. B. Shabat, Sov. Phys. JETP **34**, 62 (1972).
[8]Y. Nutku, J. Math. Phys. **26**, 1237 (1985).
[9]D. D. Holm, J. Marsden, T. Ratiu, and A. Weinstein, Phys. Rep. **123**, 1 (1985).
[10]J. Cavalcante and H. P. McKean, Physica (Utrecht) D **4**, 253 (1982).
[11]B. A. Kupershmidt, "Deformations of Hamiltonian structures," M.I.T. preprint.
[12]D. J. Benney, Stud. Appl. Math. **52**, 45 (1973).
[13]B. A. Kupershmidt, Commun. Math. Phys. **99**, 51 (1985).
[14]S. D. Poisson, J. Ec. Polytech. **7**, (1808).
[15]P. J. Morrison, "Poisson brackets for fluids and plasmas," in *Mathematical Methods in Hyrodynamics and Integrability in Dynamical Systems*, edited by M. Tabor and Y. K. Treve, AIP Conference Proceedings No. 88, 1982 (AIP, New York, 1982).
[16]J. Allen and J. Andrews, J. Plasma Phys. **4**, 187 (1970).
[17]F. Magri, C. Morosi, and O. Ragnisco, Commun. Math Phys. **99**, 115 (1985).
[18]F. Magri and C. Morosi, "A geometrical characterization of integrable Hamiltonian systems through the theory of Poisson-Nijenhuis manifolds," preprint, Universita degli Studi di Milano, 1984.
[19]P. J. Olver, *Applications of Lie Groups to Differential Equations* (Springer, Berlin, 1986).
[20]H. Flanders, *Differential Forms* (Academic, New York, 1963).
[21]B. Schutz, *Geometrical Methods of Mathematical Physics* (Cambridge U. P., Cambridge, 1980).
[22]Y. Nutku, J. Phys. A **16**, 4195 (1983).
[23]E. Suhubi, Int. J. Eng. Sci. **22**, 110 (1984).

A search for bilinear equations passing Hirota's three-soliton condition. III. Sine–Gordon-type bilinear equations

Jarmo Hietarinta
Department of Physical Sciences, University of Turku, 20500 Turku, Finland

(Received 3 March 1987; accepted for publication 10 June 1987)

In this paper the results of a search for pairs of bilinear equations of the type $A^i(D_x,D_t)F\cdot F + B^i(D_x,D_t)G\cdot F + C^i(D_x,D_t)G\cdot G = 0$, $i = 1,2$, which have standard type three-soliton solutions, are presented. The freedom to rotate in (F,G) space is fixed by the one-soliton ansatz $F = 1$, $G = e^n$, then the B^i determine the dispersion manifold while A^i and C^i are auxiliary functions. In this paper it is assumed that B^1 and B^2 are even and proportional, and that A^i and C^i are quadratic. As new results, $B^1 = aD_x^3 D_t + D_tD_y + b$, $A^2 = -C^2 = D_xD_t$, and generalizations of the sine–Gordon model $B^1 = D_xD_t + a$ with a family of auxiliary functions A^i and C^i are obtained.

I. INTRODUCTION

In recent papers we have searched for bilinear equations for which single-soliton solutions can be combined to form two- and three-soliton solutions. In Ref. 1 we studied equations of the type

$$P(D_x,D_t)f\cdot f = 0, \tag{1}$$

and in Ref. 2 pairs of bilinear equations of the type

$$P_i(D_x,D_t)F\cdot G = 0, \quad i = 1,2, \tag{2}$$

where

$$P_i(-X,-T) = (-1)^i P(X,T), \quad i = 1,2. \tag{3}$$

In this paper we generalize (2) to

$$A^i(D_x,D_t)F\cdot F + B^i(D_x,D_t)G\cdot F + C^i(D_x,D_t)G\cdot G = 0, \quad i = 1,2. \tag{4}$$

A good example of an equation that can be put in the form (4) but not in the form (2) or (3) is the sine–Gordon equation

$$u_{xt} = \sin u. \tag{5}$$

Using the dependent variable transformation[3,4]

$$u = 4\arctan(G/F) \tag{6}$$

one finds that (5) is satisfied if

$$(D_xD_t - 1)G\cdot F = 0, \quad D_xD_t(F\cdot F - G\cdot G) = 0. \tag{7}$$

In Eq. (12) of Ref. 2 we gave a rotated form of (7). Indeed, we always have various degrees of freedom that should be fixed to make computations and classifications simpler. In Ref. 2 the freedom of linear transformations $(F,G)\to(f,g)$ by

$$F = af + bf, \quad G = cf + dg, \quad ad - bc \neq 0 \tag{8}$$

(a, b, c, and d constants) was fixed by the fact that only $F\cdot G$ was to appear in (2); then it followed that the one-soliton solution (1SS) was of the type

$$F = 1 + e^n, \quad G = 1 - e^n, \tag{9}$$

$$n = px + \Omega t + m. \tag{10}$$

In this paper we fix the rotational freedom (8) by requiring the 1SS to be of the type

$$F = 1, \quad G = e^n. \tag{11}$$

This choice will simplify calculations. When (2) is rotated to allow 1SS of this type it reads

$$P_1(D_x,D_t)G\cdot F = 0, \quad P_2(D_x,D_t)(F\cdot F - G\cdot G) = 0. \tag{12}$$

Note the close analogy with (7), the only difference is that in Ref. 2 P_1 was assumed to be odd in the variables [cf. (3)], while in (7) both polynomials are even.

One cannot always rotate F and G so that the one-soliton ansatz is given by (11). For it is possible that in the ansatz

$$F = a + be^n, \quad G = c + de^n, \quad ad - bc \neq 0 \tag{13}$$

some of the numbers a, b, c, and d turn out to depend on the parameters p and Ω in a way that cannot be completely eliminated by adjusting the constant m in n of (10), and by overall multiplication. [This is what happens, e.g., when the P_i's in (2) have both odd and even terms.[5]]

In the next section we will study the conditions for A^i, B^i, and C^i following from the existence of one-, two-, three-, and four-soliton solutions. In the cases studied in Refs. 1 and 2 the existence of two-soliton solutions (2SS) was automatic, but in the present more general setting we also get more conditions. After the general discussion we report the results of our search for systems with three-soliton solutions (3SS) for the case when the B^i's are proportional and even (the odd case was discussed in Ref. 2), and A^i and C^i are quadratic.

II. THE *N*-SOLITON CONDITION

The N-soliton conditions will be more complicated now that we do not at the beginning assume any special relationships between A^i and C^i in (4). We will now go through the conditions for the existence of standard form one- to four-soliton solutions.

A. The one-soliton condition

We start with Eqs. (4), where the operators D_x and D_t are defined, as usual, by

$$D_x^{\,n}D_t^{\,m}F\cdot G = (\partial_x - \partial_{x'})^n(\partial_t - \partial_{t'})^m \times F(x,t)G(x',t')|_{x'=x,\,t'=t}. \tag{14}$$

Because $D_x{}^n D_t{}^m F\cdot G = (-1)^{n+m} D_x{}^n D_t{}^m G\cdot F$ we may assume that A^i and C^i in (4) are both even.

We take the 1SS (11), with (10) and when it is substituted into (4) we find that the equations are satisfied, if

$$A^i(0) = 0, \quad C^i(0) = 0, \tag{15}$$

$$B^i(p,\Omega) = 0, \tag{16}$$

for $i = 1,2$. Equation (16) yields the *dispersion relation*. In general let us define the affine manifold $V(P)$ related to a polynomial P by

$$V(P) = \{(x,t) | P(x,t) = 0\}, \tag{17}$$

then the result can be expressed as follows.

Theorem 1: The pair of equations (4) has 1SS's of type (11) for those parameter values (p,Ω) that belong to V_{1S}, where

$$V_{1S} = V(B^1) \cap V(B^2) \tag{18}$$

is the *dispersion manifold*.

If the B^i's are proportional we get the standard result $V_{1S} = V(B)$. In the general case the B^i's may be different. If we are working in two dimensions then each $V(B^i)$ would be the union of irreducible algebraic curves, and V_{1S} would consist of those algebraic curves corresponding to the greatest common factor of B^1 and B^2, and possible isolated points. For later use let us define here the Cartesian product of affine manifolds:

$$V_{D,n} = \{(x_1,t_1,...,x_n,t_n) | (x_i,t_i) \in V_D, \ i = 1,...,n\}. \tag{19}$$

B. The two-soliton condition

For the 2SS we take the natural generalization

$$F = 1 + K_{12} e^{n_1 + n_2}, \quad G = e^{n_1} + e^{n_2}, \tag{20}$$

where the n_i's are constructed as before in (10) with both pairs of parameters (p_i,Ω_i) in V_{1S}. When (20) is substituted into (4) we find the following additional conditions:

$$K_{12} A^i(p_1 + p_2, \Omega_1 + \Omega_2) + C^i(p_1 - p_2, \Omega_1 - \Omega_2) = 0, \quad i = 1,2, \tag{21}$$

$$K_{12} B^i(-p_j, -\Omega_j) = 0, \quad i,j = 1,2. \tag{22}$$

Let us first discuss (21). In principle it defines K_{12}, except when the solitons *resonate*, i.e., if $A^i(p_1 + p_2, \Omega_1 + \Omega_2) = 0$ for $i = 1,2$. Here we exclude this possibility and assume that $A^2(p_1 + p_2, \Omega_1 + \Omega_2)$ does not vanish. Then

$$K_{12} = -C^2(p_1 - p_2, \Omega_1 - \Omega_2)/A^2(p_1 + p_2, \Omega_1 + \Omega_2), \tag{23}$$

and the other equation ($i = 1$) yields the further condition

$$A^1(p_1 + p_2, \Omega_1 + \Omega_2) C^2(p_1 - p_2, \Omega_1 - \Omega_2) = A^2(p_1 + p_2, \Omega_1 + \Omega_2) C^1(p_1 - p_2, \Omega_1 - \Omega_2). \tag{24}$$

This is a *compatibility condition* between the auxiliary polynomials of the two equations. It implies that K_{12} in (23) is in fact independent of the superscript (as long as the rhs is not 0/0). Since the parameters (p_i,Ω_i) are in V_{1S} it is sufficient that this equation holds on $V_{1S,2}$.

If we assume that (24) is an *identity* then it is easy to see that

$$A^1 = \mu A^2, \quad C^1 = \mu C^2, \tag{25}$$

and then by taking suitable linear combinations of the equations we may take $A^1 \equiv C^1 \equiv 0$. We observe that this is indeed what happens in (12) and (7). [Note that taking linear combinations does not change V_{1S}.]

However, (24) can have additional solutions in some dispersion manifolds. For example if $B^1 = XT - 1$, $B^2 = 0$ (the sine–Gordon case) then (24) is also satisfied, e.g., if

$$C^i(X,T) = \mu A^i(X, -T), \tag{26a}$$

$$A^i(kX,kT) = k^m A^i(X,T), \tag{26b}$$

where the degree of homogeneity m must be independent of the subscript i. In fact, for this choice of A^i, B^i, and C^i, K_{ij} depends no more on the precise form of A^i and C^i, but only on m

$$K_{ij} = -\mu(p_i - p_j)^m/(p_i + p_j)^m. \tag{26c}$$

We will return to this possibility later in Sec. III C.

Next let us consider (22). In the special case where $K_{12} = 0$ this is an identity, but in the generic case (22) need not be a consequence of (16). The point is that B^i may contain both odd and even parts. Let us define

$$B^{i,e}(X,T) = [B^i(X,T) + B^i(-X,-T)]/2,$$
$$B^{i,o}(X,T) = [B^i(X,T) - B^i(-X,-T)]/2, \tag{27}$$

then (16) and (22) together mean that a 2SS is possible only if $(p_i,\Omega_i) \in V_{2S}$, where

$$V_{2S} = V(B^{1,e}) \cap V(B^{1,o}) \cap V(B^{2,e}) \cap V(B^{2,o}). \tag{28}$$

Clearly $V_{2S} \subseteq V_{1S}$. If there are pairs of parameters that belong to V_{1S} but not to V_{2S} then one cannot combine the corresponding 1SS's to a 2SS of the type (20). This would suggest that the system is not integrable and therefore a reasonable requirement for the B's is that $V_{2S} = V_{1S}$. Thus we obtain the following theorem.

Theorem 2: A pair of 1SS's of (4) as given in Theorem 1 can be combined to a 2SS (20) if (1) they do not resonate, i.e., $A^2(p_1 + p_2, \Omega_1 + \Omega_2) \neq 0$, (2) $V_{2S} = V_{1S}$ ($= V_S$), and (3) the auxiliary functions are compatible on $V_{S,2}$ in the sense of (24). The coefficient K_{12} is given by (23).

Let us observe one more property about V_{2S}. Since it is defined by polynomials of definite parity it follows that if $(p,\Omega) \in V_{2S}$ then also $(-p, -\Omega) \in V_{2S}$. This implies, e.g., that if some polynomial $P(p_i,\Omega_i,p_j,\Omega_j)$ vanishes on $V_{2S,2}$ then so does $P(p_i,\Omega_i, -p_j, -\Omega_j)$.

C. The three-soliton condition

For the 3SS we generalize (20) further to

$$F = 1 + K_{12} e^{n_1 + n_2} + K_{13} e^{n_1 + n_3} + K_{23} e^{n_2 + n_3},$$
$$G = e^{n_1} + e^{n_2} + e^{n_3} + K_{123} e^{n_1 + n_2 + n_3}, \tag{29}$$

where the n_i's are as in (10) with $(p_i,\Omega_i) \in V_S$, and

$$K_{ij} = -C^2(p_i - p_j, \Omega_i - \Omega_j)/A^2(p_i + p_j, \Omega_i + \Omega_j). \tag{30}$$

When (29) and (30) are substituted into (4) we get the following additional conditions:

$$K_{ik}K_{jk}A^m(p_i-p_j,\Omega_i-\Omega_j)$$
$$+K_{123}C^m(p_i+p_j,\Omega_i+\Omega_j)=0, \qquad (31)$$

for $m=1,2$ and for the three cases where $\{i<j,k\}$ is a permutation of $\{1,2,3\}$, and

$$K_{123}B^m(p_1+p_2+p_3,\Omega_1+\Omega_2+\Omega_3)$$
$$+K_{12}B^m(-p_1-p_2+p_3,-\Omega_1-\Omega_2+\Omega_3)$$
$$+K_{13}B^m(-p_1+p_2-p_3,-\Omega_1+\Omega_2-\Omega_3)$$
$$+K_{23}B^m(p_1-p_2-p_3,\Omega_1-\Omega_2-\Omega_3)=0, \qquad (32)$$

for $m=1,2$.

Let us first analyze (31). We assume that $C^2(p_i-p_j,\Omega_i-\Omega_j)$ does not vanish anywhere on $V_{S,2}$, and then if we denote

$$L_{ij}=\frac{A^2(p_i-p_j,\Omega_i-\Omega_j)A^2(p_i+p_j,\Omega_i+\Omega_j)}{C^2(p_i-p_j,\Omega_i-\Omega_j)C^2(p_i+p_j,\Omega_i+\Omega_j)}, \qquad (33)$$

we can write (31) for $m=2$ as

$$K_{123}=K_{12}K_{13}K_{23}L_{ij} \qquad (34)$$

where $i<j\in\{1,2,3\}$. This means that

$$L_{12}=L_{13}=L_{23} \quad (=L). \qquad (35)$$

Since there is no variable that appears in all of the L_{ij}'s we find that L is a constant on $V_{S,2}$. Thus we get the following *compatibility condition between A^2 and C^2*:

$$A^2(p_i-p_j,\Omega_i-\Omega_j)A^2(p_i+p_j,\Omega_i+\Omega_j)$$
$$=L\cdot C^2(p_i-p_j,\Omega_i-\Omega_j)C^2(p_i+p_j,\Omega_i+\Omega_j) \qquad (36)$$

on $V_{S,2}$ for some constant L.

When (34) is substituted into (31) for $m=1$, and we use (36), we get a condition similar to (24), only signs change:

$$A^1(p_i-p_j,\Omega_i-\Omega_j)C^2(p_i+p_j,\Omega_i+\Omega_j)$$
$$=A^2(p_i-p_j,\Omega_i-\Omega_j)C^1(p_i+p_j,\Omega_i+\Omega_j). \qquad (37)$$

Due to the form of V_{2S} (37) is then a consequence of (24), as was discussed at the end of Sec. II B.

Let us now consider some possible solutions of (36). If we again require it to hold identically then we must take

$$C^2=\alpha A^2. \qquad (38)$$

The constant α can be absorbed into G by changing the constant m in (10); for convenience we use this freedom to put $\alpha=-1$.

As before it is sufficient that (36) is satisfied on $V_{S,2}$. To continue with the example above ($B^1=XT-1, B^2=0$) we note that (36) holds on $V_{S,2}$ without additional assumptions for the choice (26), and $L=\mu^{-2}$.

Condition (32) is the three-soliton condition proper. When we substitute K_{ij} from (30) and K_{123} from (34) and multiply out the denominators, we can write (32) as

$$\sum_{i<j,k}^{(3)} A^2(p_i+p_k,\Omega_i+\Omega_k)A^2(p_j+p_k,\Omega_j+\Omega_k)C^2(p_i-p_j,\Omega_i-\Omega_j)B^m(-p_i-p_j+p_k,-\Omega_i-\Omega_j+\Omega_k)$$
$$+L\cdot C^2(p_1-p_2,\Omega_1-\Omega_2)C^2(p_1-p_3,\Omega_1-\Omega_3)C^2(p_2-p_3,\Omega_2-\Omega_3)B^m(p_1+p_2+p_3,\Omega_1+\Omega_2+\Omega_3)=0 \qquad (39)$$

on the dispersion manifold $V_{S,3}$. Thus we obtain the following theorem.

Theorem 3: The pair of equations (4) has 3SS solutions of type (29) if (1) the conditions of Theorem 2 are satisfied, (2) the functions A^2,C^2 are compatible on $V_{S,2}$ in the sense of (36), and (3) Eq. (39) holds on $V_{S,3}$. Then $K_{123}=LK_{12}K_{13}K_{23}$, where L is a constant defined in (36).

Note that if, following (38), we take $C^2=-A^2$ then (39) can also be written as

$$\sum_{\sigma=\pm1} B^m\left(\sigma_1\sigma_2\sigma_3\sum_{i=1}^{3}\sigma_i p_i,\sigma_1\sigma_2\sigma_3\sum_{i=1}^{3}\sigma_i\Omega_i\right)$$
$$\times\prod_{i>j}^{(3)}A^2(\sigma_i p_i-\sigma_j p_j,\sigma_i\Omega_i-\sigma_j\Omega_j)=0. \qquad (40)$$

If furthermore $B^2=0$ and B^1 is odd then this becomes Eq. (24) of Ref. 2 for $n=3$.

D. The four-soliton condition

Finally let us consider the four-soliton condition. In view of the previous results the ansatz (29) is generalized to

$$F=1+\sum_{i<j}^{(4)}K_{ij}\,e^{n_i+n_j}+Me^{n_1+n_2+n_3+n_4},$$
$$G=\sum_{i=1}^{4}e^{n_i}+\sum_{i<j<k}^{(4)}LK_{ij}\,K_{ik}K_{jk}e^{n_i+n_j+n_k}, \qquad (41)$$

where n_i, K_{ij}, and L are defined as before, and M is the new unknown. When this is substituted into (4) we get two new equations. If for M we take the expected form

$$M=L^2K_{12}K_{13}K_{14}K_{23}K_{24}K_{34}, \qquad (42)$$

then one of the equations reduces to (32) with opposite signs in the arguments of B. Because of the definite parity of the dispersion manifold this sign change is irrelevant, as was discussed at the end of Sec. II B. The other equation is

$$\sum_{i<j,k<l}^{(4)}\frac{1}{2}K_{ij}K_{kl}A(p_i+p_j-p_k-p_l,\Omega_i+\Omega_j-\Omega_k-\Omega_l)$$
$$+\sum_{i<j<k}^{(4)}LK_{ij}\,K_{ik}K_{jk}$$
$$\times C(-p_i-p_j-p_k+p_l,-\Omega_i-\Omega_j-\Omega_k+\Omega_l)$$
$$+MA(p_1+p_2+p_3+p_4,\Omega_1+\Omega_2+\Omega_3+\Omega_4)=0, \qquad (43)$$

on $V_{S,4}$. Here i,j,k,l is a permutation of $\{1,2,3,4\}$. Using (30) and (42) this equation can be written in the form

$$\sum_{\sigma=\pm1}P\left(\sigma_1\sigma_2\sigma_3\sigma_4;\sum_{i=1}^{4}\sigma_i p_i,\sum_{i=1}^{4}\sigma_i\Omega_i\right)\sigma_1\sigma_2\sigma_3\sigma_4$$
$$\times\prod_{i>j}^{(4)}P(-\sigma_i\sigma_j;\sigma_i p_i-\sigma_j p_j,\sigma_i\Omega_i-\sigma_j\Omega_j)=0 \qquad (44a)$$

on $V_{S,4}$, where we use the definition

$$P(+;X,T)=A^i(X,T),\quad P(-;X,T)=L^{1/2}C^i(X,T). \tag{44b}$$

For $C^i=-A^i, L=1$ (44a) simplifies to Eq. (25) of Ref. 2 for $n=4$. Equation (44a) is identically satisfied if A^i and C^i are proportional and quadratic.

Theorem 4: The pair of equations (4) has 4SS's of type (41) if (1) the assumptions of Theorem 3 are satisfied, and (2) Eqs. (44) hold on $V_{S,4}$. The coefficient M is given by (42).

E. Implementing the dispersion relation

The equations that we are studying [(24), (36), (39), and (44)] need to hold only when the parameters (p_i,Ω_i) belong to the dispersion manifold V_S. Thus we have to implement this fact in a consistent manner. In this paper we do not deal with the general case but assume that

$A\,1$: $B^2=cB^1$, B^1 is nonzero and has definite parity.

From this it follows immediately that

$$V_{S,n}=\{(x,t)\in\mathbb{C}^{2n}|B^1(x_i,t_i)=0,\ \forall i=1,...,n\}. \tag{45}$$

In the following we use the same methods as in Ref. 1. The polynomial B^1 is factored as

$$B^1(X,T)=\prod_j Q_j(X,T)^{n_j}, \tag{46}$$

where each Q_j is a monic irreducible polynomial. For the purpose of classification we group the irreducible factors according to their multiplicity as

$$B^1(X,T)=\prod_{i=1}^{s} B^1{}_i(X,T)^{n_i},\quad n_i>n_j \quad \text{for } i>j. \tag{47}$$

In the tables square brackets are used to separate the $B^1{}_i$'s. Next we introduce the definition

$$\sqrt{B^1}(X,T)=\prod_j Q_j(X,T)=\prod_i B^1{}_i(X,T). \tag{48}$$

The property that some polynomial $S(p_1,\Omega_1,...,p_n,\Omega_n)$ vanishes on the dispersion manifold determined by B^1 means that it should vanish on $V_{S,n}$. This, according to the theorem in Sec. III A of Ref. 1, in turn means that one can find polynomials $K_{n,i}$ in the variables $X_1,T_1,...,X_n,T_n$, so that S can be expressed as

$$S=\prod_{i=1}^{n}\sqrt{B^1}(X_i,T_i)K_{n,i}(X_1,T_1,...,X_n,T_n). \tag{49}$$

To implement (49) for computer algebra systems we introduce a consistent ordering in the set of monomials X^mT^n (for example, first by $m+n$ and then by m among those with the same $m+n$). Then for each $i=1,2,3$ we take the leading monomial of $\sqrt{B^1}(X_i,T_i)$, let us call it $\sqrt{M}(X_i,T_i)$, and replace it everywhere in S by $\sqrt{M}(X_i,T_i)-\sqrt{B^1}(X_i,T_i)$. It is easy to see that S vanishes under this rewriting rule iff it can be written as in (49). Note that the rewriting rule decreases the order of S and therefore eventually $\sqrt{M}$'s can no longer be extracted and the procedure terminates. In REDUCE the rewriting rule is accomplished by a LET statement.[6] For the LET statement it is important that the polynomial $\sqrt{M}$ is monic.

III. RESULTS

In addition to the assumption A1 in Sec. II E we assume here that

A2: A^i, C^i are quadratic.

The search problem can be divided into two broad categories according to whether Eqs. (24) and (36) are identities or not. If they are both identities then the system can be transformed, as was shown before, into the form where $A^1=C^1=0, A^2=-C^2$. Since the B^i's are proportional we may take further linear combinations so that $B^2=0$. Now that A^2 is quadratic we use linear transformations in (X,T) to take A^2 either to X^2 or to XT. Equations with odd B^1 and these auxiliary functions were searched in Ref. 2. The even B^1 case is discussed in Secs. III A and III B below.

It was shown before that Eqs. (24) and (36) may also have solutions that hold on certain dispersion manifolds. These special cases are discussed further in Sec. III C.

A. $A^2=-C^2=X^2, A^1=C^1=0$

In the search of polynomials B^1 that pass the three-soliton conditions (3SC) we follow the method described in Ref. 2. The X variable is fixed by A^2, while the T variable is defined by the highest-order factor of B^1 differing from X. We start with the leading monomial $B^1=X^MT^N$, $\sqrt{B^1}=X^KT^L$ and first find the relationship between M, N, K, and L for (40) to hold. We studied this for B^1 with total degree up to 20, and conjecture that the pattern that emerged hold for arbitrary degree. The results are as follows:

1.1: $B^1=X^{2N}$, $\sqrt{B^1}=X^K$,
$0<K\leqslant 2[(N-1)/3]+4$;

1.2: $B^1=X^{2N+1}T^{2M+1}$, $\sqrt{B^1}=XT$;

1.3: $B^1=T^{2M}$, $\sqrt{B^1}=T^K$,
$0<K\leqslant[(2M-1)/3]+1$.

(Here the square brackets stand for the integer part.) Note, e.g., that $X^{2M}T^{2N}$ is not acceptable when M and N are both nonzero. It is interesting to compare these results with those for odd B^1 in Sec. III A of Ref. 2.

The subsequent classification is described in Table I. In the first column we give the type (N,M). In the second column we have partitioned X^NT^M into those possible combinations that satisfy (47), (48), and 1.1–1.3 above. After this we considered the acceptable homogeneous generalizations of the given monomial parts. For types $(N,0)$ no such generalizations are possible, because by redefining one of the factors to T the type would become $(N-M,M)$ for some $M>0$. Also no T factors can be generalized to $T+aX$, because X is before T in our ordering of monomials. Case 1.2 could in principle generalize to $B^1=(X+aT)^{2N+1}\times T^{2M+1}$, $\sqrt{B^1}=(X+aT)T$, but it does not pass the 3SC.

In the third column we have given the possible nonhomogeneous generalizations. Note that, e.g., $[X^2]^m$ must be

TABLE I. Classification of B^1 for $A^2 = -C^2 = X^2$.

Type	Leading monomial	Possible generalizations	Accepted final result
(2.0)			
	$[X]^2$	$\cdots$	$[X]^2$
	$[X^2]$	$[X^2+1]$	$[X^2+1]$
(1.1)			
	$[XT]$	$[XT+a]$	$[XT+a]$
(0.2)			
	$[T]^2$	$\cdots$	$[T]^2$
(4.0)			
	$[X]^4$	$\cdots$	$[X]^4$
	$[X^2]^2$	$[X^2-1]^2$	$[X^2-1]^2$
	$[X]^2[X^2]$	$[X]^2[X^2+1]$	$[X]^2[X^2+1]$
	$[X^4]$	$[X^4+R_2+R_0]$	$[X^4+XT+1]$
(3.1)			
	$[X]^3[T]$	$\cdots$	$[X]^3[T]$
(2.2)			
(1.3)			
	$[X][T]^3$	$\cdots$	$[X][T]^3$
(0.4)			
	$[T]^4$	$\cdots$	$[T]^4$
	$[T^2]^2$	$[T^2+1]^2$	$\cdots$
(6.0)			
	$[X]^6$	$\cdots$	$[X]^6$
	$[X]^4[X^2]$	$[X]^4[X^2-1]$	$[X]^4[X^2-1]$
	$[X^2]^3$	$[X^2-1]^3$	$[X^2-1]^3$
	$[X]^3[X^3]$	$[X]^3[X^3+R_1]$	$\cdots$
	$[X^3]^2$	$[X^3+R_1]^2$	$[X^3+X]^2$
	$[X^2]^2[X^2]$	$[X^2-1]^2[X^2+a]$	$[X^2-1]^2[X^2+a]$
(5.1)			
	$[X]^5[T]$	$\cdots$	$[X]^5[T]$
(4.2)			
(3.3)			
	$[X]^3[T]^3$	$\cdots$	$[X]^3[T]^3$
(2.4)			
(1.5)			
	$[X][T]^6$	$\cdots$	$[X][T]^6$
(0.6)			
	$[T]^6$	$\cdots$	$[T]^6$
	$[T^2]^3$	$[T^2-1]^3$	$\cdots$

generalized as $[X^2+a]^m$ with a nonzero a, otherwise it would violate the definition of $\sqrt{B^1}$. Since the constant a is nonzero X can be scaled so that the entry becomes $[X^2+1]^m$. In column 4 of Table I we have given the results that pass the 3SC. They can be grouped as follows.

(i) We have only one genuinely nonlinear result, namely,

$$B^1 = aX^4 + XT + b, \quad A^2 = -C^2 = X^2. \tag{50}$$

Here a and b are constants which can be scaled to 1 if they are nonzero.

(ii) As for the results for which V_S splits into linear submanifolds, we have

1.A: B^1 is a polynomial in the X variable only (up to degree 6 any polynomial is acceptable if it satisfies 1.1 above);

1.B: $B^1 = X^{2N+1}T^{2M+1}$;

1.C: $B^1 = T^{2M}$.

B. $A^2 = -C^2 = XT, A^1 = C^1 = 0$

In this case the rotational degrees of freedom are fixed by the form of A^2, and what we have left is the freedom to exchange X and T and to scale them. The results for the leading monomials are as follows:

2.1: $B^1 = X^{2N}$, $\sqrt{B^1} = X^K$,

$0 < K \leqslant [(2N-1)/3] + 3$;

2.2: $B^1 = X^{2N+1}T$, $\sqrt{B^1} = X^K T$, $0 < K \leqslant 3$;

2.3: $B^1 = X^{2N+1}T^{2M+1}$, $N \geqslant M \geqslant 0$,

$\sqrt{B^1} = X^K T^L$, $K + L \leqslant 3$, $K > 0$, $L > 0$;

2.4: $B^1 = X^{2N}T^{2M}$, $N \geqslant M > 0$, $\sqrt{B^1} = XT$.

Also these results fit into the pattern of Ref. 2, Sec. III B.

In Table II we give in column 1 the type of the leading monomial and in column 2 the possible partitionings subject to the above monomial results. Since there is less freedom to transform the systems we have a higher number of distinct homogeneous generalizations in column 3. Of these the acceptable ones are given in column 4. If there is a free constant we have scaled it to 1 (or -1) if it was assumed to be nonzero and included separately the case where the constant is set to zero. Then in column 5 we give the nonhomogeneous generalizations that should be tried, and in column 6 the acceptable results. In column 7 we have finally given the possible three-dimensional generalizations.

(iii) The nonlinear results can be combined as follows:

$$B^1 = X^2 - T^2 + a, \; A^2 = -C^2 = XT, \tag{51}$$

and

$$B^1 = aX^3T + TY + b, \quad A^2 = -C^2 = XT. \tag{52}$$

Note that (52) contains (50) as a special case obtained by the substitution $T \to X$, $Y \to T$.

(iv) The linear manifold results are

2.A: any polynomial of X, up to degree 6 at least, subject to 2.1 above;

2.B: $(X - aT)^{2N}$ (in fact this can be generalized to Y^{2N});

2.C: $(X-T)^{2N+1}(X+T)^{2M+1}$;

2.D: $X^N T^M$;

2.E: $(X - aT)^{2N+1}X$, which generalizes to $Y^{2N+1}X$;

2.F: $(X^2-1)^N X^{2M+1} T$.

C. Other possibilities

In this subsection we assume only that A^i and C^i are quadratic and find out those solutions of (24) and (26) that cannot be put in the form studied in Secs. II A and II B. We assume A^2 and C^2 are not identically zero.

We take A^2 and C^2 to be arbitrary quadratic polynomials,

$$\begin{aligned} A^2 &= a_1X^2 + a_2XT + a_3T^2, \\ C^2 &= c_1X^2 + c_2XT + x_3T^2. \end{aligned} \tag{53}$$

When they are substituted in the compatibility condition (36) we find from the coefficients of $X_1^2X_2^2$, $X_1^2T_2^2$, and $T_1^2T_2^2$ that A^2 and C^2 are proportional, unless the rewrite rules connect these kinds of terms. Since the subscripts do not change in our rewrite rules it means that only in the case

TABLE II. Classification of B^1 for $A^2 = -C^2 = XT$.

Type	Leading monomial	Possible homog. generalization	Allowed homog. generalizations	Possible nonhomog. generalizations	Allowed nonhomog. generalizations	Generalizations with Y
(2.0)						
	$[X]^2$	$[X-aT]^2$	$[X]^2$	$\cdots$	$[X]^2$	
			$[X-T]^2$	$\cdots$	$[X-T]^2$	$[Y]^2$
	$[X^2]$	$[X^2+aXT+bT^2]$	$[X^2]$	$[X^2+a]$	$[X^2+a]$	
			$[X(X-T)]$	$[X(X-T)+a]$	$[X(X-T)+a]$	$[XY+a]$
			$[X^2-T^2]$	$[X^2-T^2+a]$	$[X^2-T^2+a]$	
(1.1)						
	$[XT]$	$\cdots$	$[XT]$	$[XT+a]$	$[XT+a]$	
(4.0)						
	$[X]^4$	$[X-aT]^4$	$[X]^4$	$\cdots$	$[X]^4$	
			$[X-T]^4$	$\cdots$	$[X-T]^4$	$[Y]^4$
	$[X]^3[X]$	$[X]^3[X-T]$	$\cdots$			
		$[X-T]^3[X-aT]$	$[X-T]^3[X]$	$\cdots$	$[X-T]^3[X]$	$[Y]^3[X]$
	$[X^2]^2$	$[X^2+bXT+cT^2]^2$	$[X^2]^2$	$[X^2+1]^2$	$[X^2+1]^2$	
			$[(X-T)^2]^2$	$[(X-T)^2+1]^2$	$\cdots$	
	$[X]^2[X^2]$	$[X]^2[X^2+aXT+bT^2]$	$[X]^2[X^2]$	$[X]^2[X^2+1]$	$[X]^2[X^2+1]$	
		$[X-T]^2[X^2+aXT+bT^2]$	$\cdots$			
	$[X^4]$	$[X^4+\cdots]$	$[X^4]$	$[X^4+R_2+R_0]$	$[X^4+aX^2+1]$	
(3.1)						
	$[X]^3[T]$	$\cdots$	$[X]^3[T]$	$\cdots$	$[X]^3[T]$	
	$[X]^2[XT]$	$[X]^2[(X-aT)T]$	$[X]^2[XT]$	$[X]^2[XT-1]$	$\cdots$	
		$[X-T]^2[XT]$	$\cdots$			
	$[X^3T]$	$[(X^3+\cdots)T]$	$[X^3T]$	$[X^3T+R_2+R_0]$	$[X^3T+T(bX+cT)+a]$	$[X^3T+TY+a]$
(2.2)						
	$[X]^2[T]^2$	$\cdots$	$[X]^2[T]^2$	$\cdots$	$[X]^3[T]^2$	
(6.0)						
	$[X]^6$	$[X-aT]^6$	$[X]^6$	$\cdots$	$[X]^6$	
			$[X-T]^6$	$\cdots$	$[X-T]^6$	$[Y]^6$
	$[X]^5[X]$	$[X]^5[X-T]$	$\cdots$			
		$[X-T]^5[X-aT]$	$[X-T]^5[X]$	$\cdots$	$[X-T]^5[X]$	$[Y]^5[X]$
			$[X-T]^5[X+T]$	$\cdots$	$[X-T]^5[X+T]$	
	$[X]^4[X]^2$	$[X]^4[X-T]^2$	$\cdots$			
		$[X-T]^4[X-bT]^2$	$\cdots$			
	$[X]^4[X^2]$	$[X]^4[X^2+\cdots]$	$[X]^4[X^2]$	$[X]^4[X^2-1]$	$[X]^4[x^2-1]$	
		$[X-T]^4[X^2+\cdots]$	$\cdots$			
	$[X^2]^3$	$[X^2+\cdots]^3$	$[X^2]^3$	$[X^2+1]^3$	$[X^2+1]^3$	
			$[X^2-T^2]^3$	$[X^2-T^2+a]^3$	$[X-T]^3[X+T]^3$	
			$[(X-T)^2]^3$	$[(X-T)^2+a]^3$	$\cdots$	
	$[X]^3[X]^2[X]$	$[X]^3[X-T]^2[X-bT]$	$\cdots$			
		$[X-T]^3[X-bT]^2 \times [X-cT]$	$\cdots$			
	$[X]^3[X^3]$	$[X]^3[X^3+\cdots]$	$[X]^3[X^3]$	$[X]^3[X^3+R_1]$	$\cdots$	
		$[X-T]^3[X^3+\cdots]$	$\cdots$			
	$[X^3]^2$	$[X^3+\cdots]^2$	$[X^3]^2$	$[X^3+R_1]^2$	$[X^3+R_1]^2$	$[X^3+X]^2$
	$[X^2][X^2]^2$	$[X^2+\cdots][X^2+\cdots]^2$	$[X^2][X^2]^2$	$[X^2+1][X^2+a]^2$	$[X^2+1][X^2+a]^2$	
(5.1)						
	$[X]^5[T]$	$\cdots$	$[X]^5[T]$	$\cdots$	$[X]^6[T]$	
	$[X]^4[XT]$	$[X]^4[(X-aT)T]$	$[X]^4[XT]$	$[X]^4[XT+1]$	$\cdots$	
		$[X-T]^4[XT]$	$\cdots$			
	$[X]^3[X]^2[T]$	$[X]^3[X-T]^2[T]$	$\cdots$			
		$[X-T]^3[X]^2[T]$	$\cdots$			
	$[X]^3[X^2T]$	$[X]^3[(X^2+\cdots)T]$	$[X]^3[X^2T]$	$[X]^3[X^2T+R_1]$	$[X]^3[(X^2+1)T]$	
		$[X-T]^3[XT(X+aT)]$	$\cdots$			
	$[X^2]^2[XT]$	$[X^2+aXT+bT^2]^2[XT]$	$[X^2]^3[XT]$	$[X^2-1]^3[XT+a]$	$[X^2-1]^3[XT]$	
		$[X(X+aT)]^2[(X-T)T]$	$\cdots$			
(4.2)						
	$[X]^4[T]^2$		$[X]^4[T]^2$	$\cdots$	$[X]^4[T]^2$	
(3.3)						
	$[X]^3[T]^3$		$[X]^3[T]^3$	$\cdots$	$[X]^3[T]^3$	

of a quadratic B^1 may we obtain new solutions to (36). If B^1 is of higher degree than 2 it might still be possible that $C^i = \alpha^i A^i$, but that A^1 and A^2 are not proportional or $\alpha^1 \neq \alpha^2$. For this we would need a third-order B^1, however, detailed calculations show that (24) does not have such solutions. Thus to obtain new results we must assume that B^1 is quadratic; we fix the freedom of linear transformations partly by taking $B^1 = XT + a$, rather than by fixing A^2 as was done in III A and III B. (The case $B^1 = X^2 + 1$ has only resonating solutions.)

When (53) is substituted into (36) and $B^1 = XT + a$ we obtain the following relations between the coefficients:

$$c_1^2 = a_1^2, \quad c_3^2 = a_3^2, \quad c_2^2 = a_2^2 - 2(a_1 a_3 - c_1 c_3). \tag{54}$$

This result is best interpreted if A^2 and C^2 are factored: (54) implies

$$\begin{aligned} A^2 &= (\alpha_1 X + \beta_1 T)(\alpha_2 X + \beta_2 T), \\ C^2 &= \pm(\alpha_1 X \pm \beta_1 T)(\alpha_2 X \pm \beta_2 T), \end{aligned} \tag{55}$$

where α_i and β_i are arbitrary, and *all sign combinations are acceptable*. The overall sign of C^2 can always be absorbed into G; we use the $+$ sign. Depending on the choice of the other signs the result can be divided into the following three categories.

(i) If *both signs in* C^2 *are negative*, then we have $C^2(X,T) = A^2(X, -T)$, which was briefly mentioned earlier. Since now the explicit form of A^2 does not enter in K_{ij} [(26c)] there will be no restrictions on A^2 from the three-soliton condition. Next we tested which A^1 and C^1 are compatible with the chosen A^2 and C^2 [Eq. (24)] and found that the same relation $C^1(X,T) = A^1(X, -T)$ is required. Since this relationship is preserved when we take linear combinations of the equations we may assume that $B^2 = 0$. Thus every pair of bilinear equations,

$$\begin{aligned} &(d_1 D_x^2 + d_3 D_t^2)(F\cdot F + G\cdot G) + d_2 D_x D_t (F\cdot F - G\cdot G) \\ &\quad + (D_x D_t + b) G\cdot F = 0, \\ &(a_1 D_x^2 + a_3 D_t^2)(F\cdot F + G\cdot G) \\ &\quad + a_2 D_x D_t (F\cdot F - G\cdot G) = 0, \end{aligned} \tag{56}$$

has three-soliton solutions. (Here it is assumed that of the various constants at least one a_i is nonzero.) Equation (51) is a rotated form of the subcase $d_i = 0$, $a_1 = -a_3 = 1$, $a_2 = 0$, while in the standard sine–Gordon model (7) we have $d_i = 0$, $a_1 = a_3 = 0$, $a_2 = 1$.

Let us next consider the 4SC (43). Due to (26c) the explicit forms of the auxiliary functions appear in (43) linearly. Since the equation holds in the special case of monomial pairs $C^2(X,T) = A^2(X, -T)$ it holds for the general case (56). Thus Eqs. (56) have also 4SS's.

(ii) *One sign in* C^2 *is negative, say the one for* β_1, *and* α_2, *and* β_2 *are nonzero.* Then, after scaling the nonzero constants to 1 we obtain

$$\begin{aligned} A^1 &= (\alpha_1 X + \beta_1 T)(X + T), \\ C^2 &= (\alpha_1 X - \beta_1 T)(X + T). \end{aligned} \tag{57}$$

The condition that A^1 and C^1 are compatible with these [Eq. (24)] implies that A^1 and C^1 must have the same form as in (57) with possibly different constants. The overall common factor $X + T$ can in fact be elevated to a new independent variable Y, and still (24) and the 3SC are satisfied. By taking a linear combination we put $B^2 = 0$, and thus obtain the result that the pair of bilinear equations,

$$\begin{aligned} &\alpha_1 D_x D_y (F\cdot F + G\cdot G) + \beta_1 D_t D_y (F\cdot F - G\cdot G) \\ &\quad + (D_x D_t + a) G\cdot F = 0, \\ &\alpha_2 D_x D_y (F\cdot F + G\cdot G) + \beta_2 D_t D_y (F\cdot F - G\cdot G) = 0, \end{aligned} \tag{58}$$

has three-soliton solutions. As usual this property holds also for any projection $D_y \to c D_x + d D_t$. Also the 4SC (44) is satisfied for all parameter values.

(iii) *Both signs are positive and* α_i *and* β_i *are nonzero.* In this case A^2 and C^2 are obviously compatible, but they do not pass even the three-soliton condition. [It would require one a_i in (53) to be zero, but then we get one of the cases above.]

We have also tried to generalize these results to higher dimensions. If B^1 is still two dimensional, (58) is the most general result with a Y dependence. If B^1 is three dimensional we transform it into $B^1 = X^2 + T^2 + Y^2 + a$, then we find that $A^2 = -C^2 = X^2 + T^2 + Y^2$ is the only possibility, but even that is not acceptable unless the constant a in B^1 vanishes. Thus we confirm the results that the sine–Gordon model in dimensions higher than 2 have no general 3SS.[7]

IV. CONCLUSIONS

The bilinear equations studied in this paper are close to the ones studied in Ref. 2. In both we have quadratic auxiliary functions (A, C here, P_2 in Ref. 2); in this paper the dispersion equation is given by an even polynomial B^1, while in Ref. 2 P_1 was odd.

The new results that we have found are the following: For $\deg(B^1) > 2$ we obtain one new model, (52). It carries some resemblance to the shallow water-wave equations of Ref. 8. For a quadratic B^1 (transformed into $B^1 = XT + a$) we obtain some rather interesting results. In this case the auxiliary functions can be substantially generalized from the usual sine–Gordon result $A^2 = -C^2 = XT$, as was shown in Eqs. (56) and (58). In addition to these new and genuinely nonlinear results we have as before infinite sequences of models with linear dispersion manifolds.

ACKNOWLEDGMENTS

I would like to thank A. Ramani for discussions. The computations were done with IBM 3033 at the University of Turku Computing Center using REDUCE 3.2.

The author was supported by the Academy of Finland.

[1] J. Hietarinta, J. Math. Phys. **28**, 1732 (1987).
[2] J. Hietarinta, J. Math. Phys. **28**, 2094 (1987).
[3] R. Hirota, J. Phys. Soc. Jpn. **33**, 1459 (1972).
[4] R. Hirota, Prog. Theor. Phys. **52**, 1498 (1974).
[5] R. Hirota and J. Satsuma, Prog. Theor. Phys. **57**, 797 (1977).
[6] A. C. Hearn, *REDUCE User's Manual Version 3.2* (Rand, Santa Monica, 1985), Publ. CP78 (Rev. 4/85).
[7] R. Hirota, J. Phys. Soc. Jpn. **35**, 1566 (1973); J. D. Gibbon and G. Zambotti, Nuovo Cimento B **28**, 1 (1975); K. Kobayashi and M. Izutsu, J. Phys. Soc. Jpn. **41**, 1091 (1976).
[8] R. Hirota and J. Satsuma, J. Phys. Soc. Jpn. **40**, 611 (1976); M. Ito, *ibid.* **49**, 286 (1980).

Nonlinear resonance for quasilinear hyperbolic equation

Tai-Ping Liu
Department of Mathematics, University of Maryland, College Park Campus, College Park, Maryland 20742

(Received 19 March 1987; accepted for publication 3 June 1987)

The purpose of this paper is to study the wave behavior of hyperbolic conservation laws with a moving source. Resonance occurs when the speed of the source is too close to one of the characteristic speeds of the system. For the nonlinear system characteristic speeds depend on the basic dependence variables and resonance gives rise to nonlinear interactions which lead to rich wave phenomena. Motivated by physical examples a scalar model is proposed and analyzed to describe the qualitative behavior of waves for a general system in resonance with the source. Analytical understanding is used to design a numerical scheme based on the random choice method. An important physical example is transonic gas flow through a nozzle. This analysis provides a transparent and revealing qualitative understanding of wave behavior of gas flow, including such phenomena as nonlinear stability, instability, and changing types of waves.

I. INTRODUCTION

Consider a hyperbolic conservation law with a moving source

$$\frac{\partial u}{\partial t}+\frac{\partial f(u)}{\partial x}=C(x-\alpha t)h(u),$$

where α is the speed of the source. Our purpose is to study the qualitative behavior of the solution. By a change of variables $x-\alpha t\to x$ and $f(u)\to f(u)-\alpha u$, we may assume that the source is stationary,

$$\frac{\partial u}{\partial t}+\frac{\partial f(u)}{\partial x}=C(x)h(u),\quad u\in\mathbb{R}^1. \tag{1.1}$$

Several physical situations can be modeled as hyperbolic conservation laws with a moving source. For instance, a moving magnetic field for magnetohydrodynamics (MHD) and the geometric effect of a nozzle on the gas flow can be expressed as a moving source. The quasi-one-dimensional model of gas flow through a nozzle is

$$\frac{\partial \mathbf{u}}{\partial t}+\frac{\partial \mathbf{f}(\mathbf{u})}{\partial x}=C(x)\mathbf{h}(\mathbf{u}),\quad \mathbf{u}\equiv(\rho,\rho V,\rho E)^t,$$

$$\mathbf{f}(\mathbf{u})\equiv(\rho V,\rho V^2+P,\rho EV+PV)^t,$$

$$C(x)\equiv -A'(x)/A(x),$$

$$\mathbf{h}(\mathbf{u})\equiv(\rho V,\rho V^2,\rho EV+PV)^t,$$

where ρ, V, P, and E are the density, velocity, pressure, and total energy of the gas, and $A(x)$ is the cross section of the nozzle. For uniform nozzle $C(x)\equiv 0$ the system becomes the compressible Euler equations. The equations have three characteristic speeds

$$\lambda_1=V-C,\quad \lambda_2=V,\quad \lambda_3=V+C,$$

where C is the sound speed. Interesting wave phenomena occur when the flow is transonic, that is, $V-C$ changes signs. In this case λ_1 is around zero, which is the speed of the source. The shape of the nozzle has stabilizing and destabilizing effects, see Liu.[1] There are a finite number of asymptotic shapes that can be constructed explicitly. These qualitative properties hold also for the general system of hyperbolic conservation laws with a moving source, see Li–Liu.[2]

Motivated by the aforementioned studies of physical models, we make the following assumptions on the source. The strength of the source is measured by $C(x)$. We assume that the strength of the source is finite. For definiteness we suppose that $C(x)$ is piecewise smooth and

$$C(x)=0\quad \text{for } x\notin[0,1]. \tag{1.2}$$

The function $h(u)$ represents the coupling of the source with the hyperbolic conservation law. The model (1.1) is meant to study the family of waves for the general system which is in resonance with the source. For the transonic flow through a nozzle, we want to study the behavior of waves pertaining to the characteristic value λ_1. With these physical considerations in mind we make the following nondegeneracy assumption:

$$h(u)\neq 0,\quad h'(u)\neq 0, \tag{1.3}$$

and that $h(u)$ is smooth for all u under consideration. Without this crucial assumption of the strong coupling of the source with the conservation law neutrally stable waves exist and there are infinitely many asymptotic states, thereby not reflecting the strongly nonlinear nature of the physical phenomena just mentioned.

Since we are interested in the nonlinear resonance, that is, the characteristic speed $f'(u)$ of (1.1) is around zero, we assume that $f'(u)$ changes signs for the range of u under consideration. For simplicity, and with fluid motion in mind, we assume that $f'(u)$ varies monotonically with u,

$$f''(u)>0. \tag{1.4}$$

By translation and rotation we may assume, without loss of generality, that

$$f(0)=f'(0)=0. \tag{1.5}$$

From (1.4) and (1.5), $f'(u)>0$ for $u>0$. Thus the state zero is called *sonic*, and any positive state $u>0$ is called *supersonic*, negative states $u<0$ *subsonic*. A wave is transonic if it takes both supersonic and subsonic values.

For the hyperbolic conservation law associated with (1.1),

$$\frac{\partial u}{\partial t}+\frac{\partial f(u)}{\partial x}=0, \tag{1.6}$$

there are shock waves and rarefaction waves. Nonlinear behavior of these waves has been extensively studied (see Ref. 3, and references therein). For (1.1) we need to consider another family of waves, the stationary waves, which move with the same speed [assumed zero in (1.1)] as the source and satisfy

$$\frac{\partial f(u)}{\partial x}=C(x)h(u). \tag{1.7}$$

To gain the basic understanding of the behavior of general solutions of (1.1) we have to study the interaction of these three types of waves. Of particular interest is when the shock waves and rarefaction waves have speed near zero and nonlinear resonance occurs. In Sec. II we study the interaction of a transonic shock wave with a stationary wave. Simple calculations exhibit the stabilizing and destabilizing effect of the source in a transparent way. In Sec. III we investigate the propagation of a rarefaction wave through a stationary wave. It is shown that the rarefaction wave reflects as a compression wave upon reaching the sonic state. The stability criterion is the following:

$C(x)h'(u)<0$ (nonlinear stability),

$C(x)h'(u)>0$ (nonlinear instability),

for $x\in[0,1]$ and all u under consideration. These studies yield a simple and revealing qualitative understanding of rich wave behavior of the compressible fluid, including such phenomena as nonlinear stability, instability, and changing types of waves.

Although basic understanding of some of the wave phenomena for general systems has been obtained, the time evolution of general solutions has not been studied. The evolution of a general solution of (1.1) can be complex but time asymptotically a general solution tends to a simple noninteracting wave pattern. In Sec. IV we study these asymptotic wave patterns. To study the evolution of general solutions in Sec. V we employ the analysis of the preceding three sections to design a numerical scheme for (1.1). The scheme is based on the random choice method, Glimm,[4] and uses shock waves, rarefaction waves, and stationary waves as building blocks. For nontransonic flows the scheme was introduced and analyzed by Liu.[5] The basic idea of generalizing it to the transonic case is the following: Instead of following the nonlinear interaction of elementary waves, we install the asymptotic state of the local wave interactions at each next time step. Since the asymptotic state is reached only at $t=\infty$, and the wave strength behaves singularly near the sonic state, the consistency of the scheme is not obvious. Nevertheless we apply the wave tracing technique of Liu[6] to show that the scheme is consistent when the sampling sequence is sufficiently equidistributed. Finally in Sec. VI we study the behavior of the solution based on the scheme.

The results in this paper have been announced in Ref. 7. Our scheme has been generalized to the gasdynamic equations for numerical purposes.[8,9] For refinement and numerical implementation of the scheme in Ref. 10, see Refs. 5 and 11. Analytical works for models not satisfying (1.3) are carried out in Refs. 12 and 13. Numerical computation using piecewise steady elements was also used in Ref. 14. For the asymptotic approximation of nozzle flow by Chisnell and Chester, see Ref. 15.

II. NONLINEAR STABILITY AND INSTABILITY OF SHOCK WAVES

Consider the propagation of a shock wave through stationary waves. The stationary wave to the right (left) of the shock wave is denoted by $u_e(x)$ [$u_r(x)$] and the location of the shock wave is $x=x(t)$, see Fig. 1. It follows from the analysis below that such a solution of (1.1) exists.

The speed of the shock wave is governed by the jump (Rankine–Hugoniot) condition

$$x'(t)=\frac{f(u_+(t))-f(u_-(t))}{u_+(t)-u_-(t)}\equiv\sigma(u_-(t),u_+(t)),$$
$$u_+(t)\equiv u_r(x(t));\quad u_-(t)\equiv u_e(x(t)). \tag{2.1}$$

Since $u_-(t)$ and $u_+(t)$ are the values of the stationary waves $u_e(x)$ and $u_r(x)$ at $x=x(t)$ we have from the stationary equation (1.7) that along $x=x(t)$

$$\frac{du_-(t)}{dt}=\frac{du_e(x(t))}{dx}x'(t)=f'(u_-(t))^{-1}C(x(t))h(u_-(t))x'(t),$$
$$\frac{du_+(t)}{dt}=\frac{du_r(x(t))}{dx}x'(t)=f'(u_+(t))^{-1}C(x(t))h(u_+(t))x'(t). \tag{2.2}$$

Differentiate (2.1) and use (2.2) to obtain

$$\frac{x''(t)}{x'(t)}=\frac{C(u(t))}{u_+(t)-u_-(t)}\Big\{h(u_+(t))-h(u_-(t))-x'(t)\Big(\frac{h(u_+(t))}{f'(u_+(t))}-\frac{h(u_-(t))}{f'(u_-(t))}\Big)\Big\}. \tag{2.3}$$

The system of ordinary differential equations (2.1) and (2.2) determines the location $x=x(t)$ and the states $u(t)$ of the shock wave. It also shows that when a shock wave propagates through a stationary wave it leaves behind another stationary wave, which is the extension of the original stationary wave behind the shock wave. One has to check, though, that it is possible to extend the stationary wave. This is a concern because Eq. (1.7) is singular at $u=0$. Nevertheless, it causes no problem because when the shock wave moves to the right (left) and the left (right) state is close to zero then the shock wave has nonpositive(non-negative) speed and thereby it does not propagate further to the right (left).

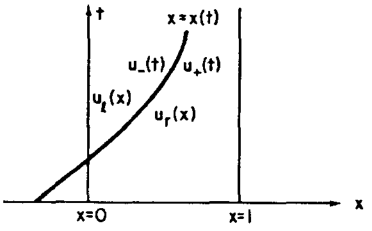

FIG. 1. Shock wave propagating through a stationary wave.

We next investigate the stability of shock waves based on (2.3). A supersonic (subsonic) shock wave, $u_- > u_+ > 0$ ($0 > u_- > u_+$), accelerates to the right (left) and moves away from the sonic state. The stability issue is relevant only for transonic shock waves $u_- > 0 > u_+$. We have from (2.3) that for nearly stationary shock waves

$$\frac{x''(t)}{x'(t)} \sim \frac{C(x(t))[h(u_+(t)) - h(u_-(t))]}{u_+(t) - u_-(t)}, \quad \text{when } x'(t) \sim 0. \tag{2.4}$$

Thus a nearly stationary transonic shock wave decelerates and is nonlinearly stable if

$$C(x)h'(u) < 0 \quad \text{(stability)} \tag{2.5a}$$

(Fig. 2), and accelerates and is nonlinearly unstable if

$$C(x)h'(u) > 0 \quad \text{(instability)} \tag{2.5b}$$

(see Fig. 3).

III. REFLECTION OF EXPANSION WAVES INTO COMPRESSION WAVES

In this section we study the propagation of a rarefaction wave through a stationary wave. Suppose that both waves are supersonic and at $t = 0$ the rarefaction wave lies in $x < 0$ propagating to the right toward the stationary wave. We also assume $C(x)$ is smooth and $C(x)h(u) < 0$ for $0 \leqslant x \leqslant 1$. With this setting the rarefaction wave may reflect as a compression wave. Other cases can be treated by the same analysis below. Denote by (u_1,u_2) the rarefaction wave and $u_0(x)$, $0 \leqslant x \leqslant 1$, the stationary wave with end states $u_2 = u_0(0)$, $u_3 = u_0(1)$ (see Fig. 4).

Since $C(x)h(u) < 0$ we have from (1.7) that $0 < u_3 < u_2$. The rarefaction wave (u_1,u_2) is an expansion wave, that is, $f'(u)$ increases across it. From the convexity assumption (1.4) we have $u_1 > u_2 > 0$. Consider the characteristic curves

$$\frac{dx}{dt} = f'(u(x,t)). \tag{3.1}$$

Smooth solutions of (1.1) satisfy

$$\frac{du}{dt} = C(x)h(u). \tag{3.2}$$

For nontransonic flow, $f'(u) \neq 0$, we may parametrize each characteristic curve by x

$$\frac{d}{dx} = \frac{dt}{dx}\frac{d}{dt} = (f'(u))^{-1}\frac{d}{dt} = (f'(u))^{-1}\left(\frac{\partial}{\partial t} + f'(u)\frac{\partial}{\partial x}\right). \tag{3.3}$$

From (3.2) and (3.3) along each characteristic curve smooth solutions satisfy

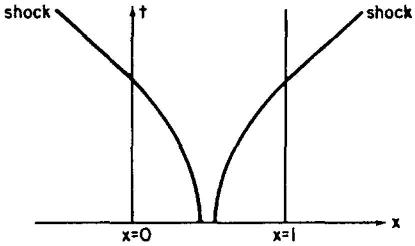

FIG. 2. Unstable shock waves, $Ch' > 0$.

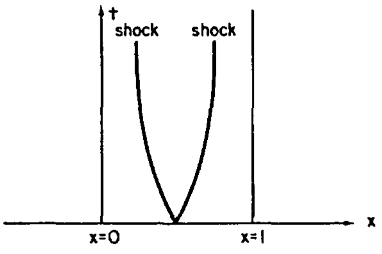

FIG. 3. Stable shock waves, $Ch' < 0$.

$$\frac{df(u)}{dx} = C(x)h(u), \tag{3.4}$$

which is of the same form as the stationary equation (1.7). From (1.1) we have

$$\frac{dw}{dt} + f''(u)(f'(u))^{-1}(C(x)h(u) - w)w = C(x)h'(u)w, \tag{3.5}$$

$$V = (f'(u))^{-1}(C(x)h(u) - w), \quad V = \frac{\partial u}{\partial x}, \quad w \equiv \frac{\partial u}{\partial t}. \tag{3.6}$$

For the supersonic flow under consideration it follows from (3.2) and the hypothesis $C(x)h(u) < 0$ that along each characteristic curve u decreases toward the sonic state zero. If the initial values of the rarefaction wave (u_1,u_2) are far from the sonic state then u may remain supersonic all along the characteristic curve. From (3.4) this is so if for any value $\bar{u}$, $u_1 \leqslant \bar{u} \leqslant u_2$, a supersonic stationary wave $u(x)$, $0 \leqslant x \leqslant 1$, exists with initial state $u(0) = \bar{u}$. By the monotonicity property of the solution operator of (3.4) this is equivalent to $u_1 \geqslant u^*$, where u^* is the unique supersonic state for which there exists a stationary wave $u(x)$ with $u(0) = u_*$ and $u(1) = 0$. Since all states are supersonic, it follows from (3.5) and (3.6) that a smooth solution exists globally. Above the characteristic curve χ with initial value u_1, w is zero at $x = 0$ and by (3.5) it remains zero for all x. Thus the solution becomes a stationary wave (u_1,u_4) followed by a rarefaction wave (u_4,u_3) after χ has passed the region $0 \leqslant x \leqslant 1$ (see Fig. 5). Here u_4 is the unique supersonic state which can be connected to u_1 by a stationary wave.

When the left state of the rarefaction wave is close to the sonic state, $u_1 < u^*$, the portion (u^*,u_2) of the rarefaction wave passes through the stationary wave as before and emerges on the other side as $(0,u_3)$. The interaction of (u,u^*) with the stationary wave needs further investigation. The characteristic curve with initial state u, $u_1 \leqslant u \leqslant u^*$ reaches the sonic state in finite time. Since $f(u)$ is quadratic at 0, (1,4) and (1,5), around the sonic state Eq. (3.1) is compared qualitatively to

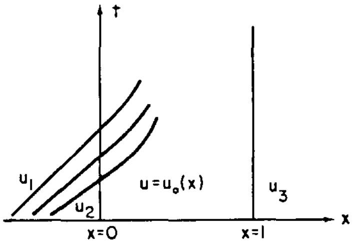

FIG. 4. Rarefaction wave propagating through a stationary wave.

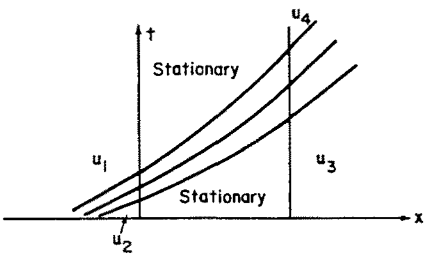

FIG. 5. $u_1 \geqslant u^*$.

$$\frac{dx}{dt} = \pm |x - x_0|^{1/2}. \tag{3.7}$$

Thus at (x_0,t_0), where the characteristic curve $x = x(t)$ becomes vertical,

$$x(t) - x_0 = |t - t_0|^2. \tag{3.8}$$

The characteristic curve with initial value u_* becomes vertical at $x = 1$. Since (3.7) is singular, it can be continued either along the vertical line $x = 1$ or backward χ_* (see Fig. 6). Other characteristic curves to the left of χ_* also turn backward after reaching the sonic state and the flow in $x < 1$ becomes subsonic.

Characteristic curves to the left of χ_* have smaller initial value and satisfy the same autonomous equation (3.4), and therefore reach sonic state at smaller x values. Consequently, between the characteristic curve χ_1 with initial value u_1 and λ_* the solution becomes subsonic compressive in finite time. Many shock waves may result from the compression. In finite time these shock waves will combine to form a transonic shock wave. In the unstable case, (2.5b), the shock wave accelerates toward $x = -\infty$ and leaves behind a subsonic stationary wave $(u_*,0)$ (see Fig. 7). Here u_* is the unique subsonic state which is connected to sonic by a stationary wave.

In the stable case, (2.5a), the transonic shock wave decelerates. Depending on the value of u_1, when u_1 is close to sonic, the speed of $\sigma(u_1,u_*)$ of the shock wave (u_1,u_*) is negative and the transonic shock wave propagates toward $x = -\infty$ (see Fig. 7). When $\sigma(u_1,u_*) > 0$, the transonic shock wave decelerates to stay in $0 < x < 1$ (see Fig. 8).

This completes the study of a rarefaction wave propagating to the right through a stationary wave. A rarefaction wave propagating to the left moves away from the sonic state and no reflection occurs. In the case $C(x)h(u) > 0$, $0 \leqslant x \leqslant 1$, the situation is just opposite, reflection occurs only for a rarefaction wave propagating to the left through a stationary wave.

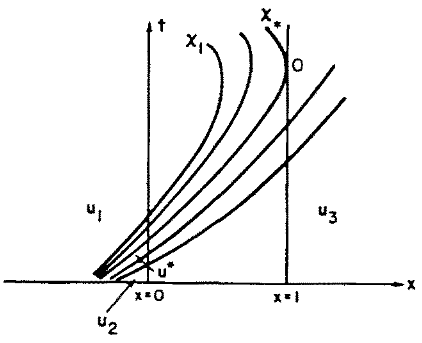

FIG. 6. Turning back of rarefaction waves.

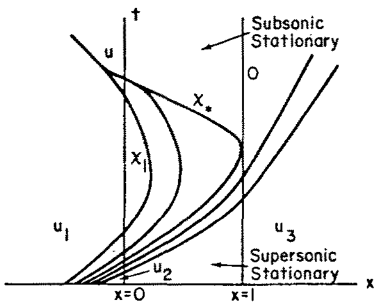

FIG. 7. Stability and $\sigma(u_1,u_*) < 0$ or instability.

IV. TIME-ASYMPTOTIC STATES

If one is not concerned with the details of the evolution process, some of the qualitative properties described in the preceding sections can be predicted easily through the study of noninteracting waves of (1.1). Thus it is composed of negative- (positive-) speeded shock waves or rarefaction waves in $x < 0$ $(x > 1)$ and a stationary wave in $0 \leqslant x < 1$. Such a wave pattern is called time asymptotic because after complex nonlinear interactions, waves are expected to combine and cancel and eventually become noninteracting (see Sec. VI). We will carry out our analysis for the case

$$C(x)h(u) < 0 \quad \text{for } 0 \leqslant x \leqslant 1. \tag{4.1}$$

The case where $C(x)h(u) > 0$, or more generally $C(x)$, changes signs can be studied by the same analysis. With (4.1), a smooth stationary wave moves away from the sonic state as x increases, (1.4), (1.5), and (1.7). Thus given a state $u \neq 0$, there exists a unique state $\tilde{u}$ such that u and $\tilde{u}$ are connected by a smooth stationary wave $u(x)$

$$u(0) = \tilde{u}, \quad u(1) = u. \tag{4.2}$$

The state u is closer to sonic than $\tilde{u}$. Since (1.4) is singular at sonic, there exist a supersonic state u^* and a subsonic state u_*; both can be connected to the sonic state in the downstream by a smooth stationary wave

$$\tilde{0} = \{u_*, u^*\}, \quad u_* < 0 < u^*. \tag{4.2'}$$

A stationary wave may be nonsmooth and contains a stationary shock wave. From (1.4) and (2.1) a given state u can be connected by a stationary shock wave to a unique state $\bar{u}$

$$f(u) \equiv f(\bar{u}), \quad u \neq \bar{u}. \tag{4.3}$$

We now construct asymptotic states connecting given states u_e on the left and u_r on the right; first for the stable case $C(x)h'(u) < 0$, $0 < x < 1$. We claim the following basic inequality:

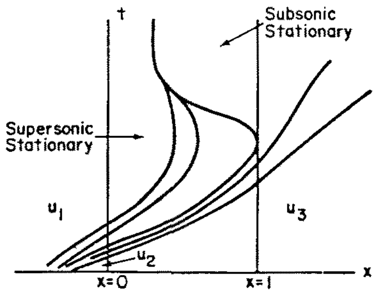

FIG. 8. Stability and $\sigma(u_1,u_*) > 0$.

$\bar{\tilde{u}} > \tilde{\bar{u}}$ for any $u>0$, when $C(x)h'(u)<0$, $0<x<1$. (4.4)

By taking $u = 0^+$ an immediate corollary is the following:

$u^* > \bar{u}_* > 0$, when $C(x)h'(u)<0$, $0<x<1$.

From (1.4) and (1.5) we see that (4.4) is equivalent to

$f(\bar{\tilde{u}}) > f(\tilde{u})$.

Let $u_1(x)$ and $u_2(x)$ be the smooth stationary waves with initial values u and $\bar{u}$, respectively. Since $u>0$, clearly $u_1(x)>0>u_2(x)$, and so $C(x)h(u_1(x)) < C(x)h(u_2(x))$ in the stable case $C(x)h'(u)<0$, $0<x<1$. We know from (4.3) that at $x=0$ $f(u_1(x)) = f(u) = f(\bar{u}) = f(u_2(x))$. Thus the solutions $u_1(x)$ and $u_2(x)$ of the stationary equation (1.7) satisfy

$f(u_2(x)) > f(u_1(x))$, $0<x\leqslant 1$.

Evaluated at $x=1$, we have $f(\bar{u}) = f(u_2(1)) > f(u_1(1)) = f(u)$. This proves (4.4). According to the relative positions of u_l and u_r, we have the following cases.

Case 1.1: $u_l \leqslant \bar{u}_*$. When u_r 0, we connect u_l to u_* by a backward wave, u_* to 0 by a subsonic stationary wave, and 0 to u_r by a forward rarefaction wave. The backward wave is a rarefaction wave (shock wave) when $u_l \leqslant u_*$ ($u_l > u_*$) (see Figs. 9 and 10).

When $u_r < 0$ the asymptotic state consists of a backward wave $(u_l \tilde{u}_r)$ and a stationary wave $(\tilde{u}_r, u_r)$.

Case 1.2: $\bar{u}_* < u_l < u^*$. When $u_r \geqslant 0$, the asymptotic state contains a supersonic rarefaction wave $(0, u_r)$. The remaining of the asymptotic state may be viewed as the resulting wave pattern of the interaction of the shock wave (u_l, u_*) in $x\leqslant 0$ and the stationary wave $(u_*, 0)$. Since $u_l > \bar{u}_*$, the speed of (u_l, u_*) is positive. As (u_l, u_*) propagates to the right through $(u_*, 0)$ it decelerates (Sec. II). Since $u_l < u_*$, the stationary wave to the left of the shock wave with initial value u_l at $x=0$ cannot be extended to the whole interval $0\leqslant x\leqslant 1$. Consequently, the shock wave eventually decelerates to become a stationary shock wave at x_0, $0<x_0<1$. Another argument, more direct and without resorting to the stability analysis in Sec. II, can also easily be put forth to conclude that the asymptotic state consists of a supersonic stationary wave for $0\leqslant x < x_0$, a stationary wave at $x = x_0$, a subsonic stationary wave for $x_0 < x \leqslant 1$, and a supersonic rarefaction wave $(0, u_r)$ for $x>0$ (see Fig. 11). Here x_0 is uniquely determined by u_l.

When $u_r < 0$ and $\bar{u}_l < \tilde{u}_r$ the asymptotic state consists of a supersonic stationary wave over $0\leqslant x < x_0$, a stationary shock x_0 is uniquely determined by u_l and u_r.

When $u_r < 0$ and $\tilde{u}_r \leqslant \bar{u}_l$, the asymptotic state consists of a backward shock wave $(u_l, \tilde{u}_r)$ and a subsonic stationary wave $(\tilde{u}_r, u_r)$.

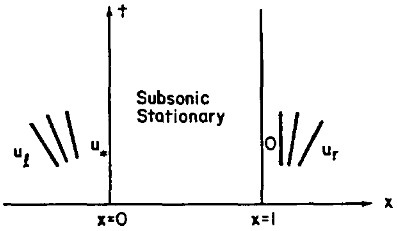

FIG. 9. $u_l \leqslant u_*$, $u_r \geqslant 0$.

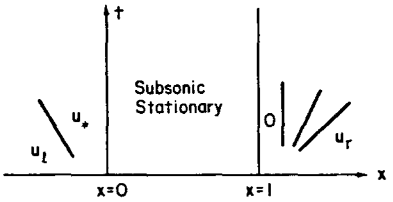

FIG. 10. $\bar{u}_* \geqslant u_l > u^*, u_r > 0$.

Case 1.3: $u_l \geqslant u^*$. Since $u_l \geqslant u^*$, there exists a supersonic state u_1 with $\tilde{u}_1 = u_l$. When $u_r \geqslant \bar{u}_1$ the asymptotic state consists of a stationary wave (u_l, u_1) and a forward wave (u_1, u_r).

When $u_r < \bar{u}_1$ and $\tilde{u}_r \geqslant \bar{u}_l$ the asymptotic state consists of transonic stationary waves with a stationary shock wave at $x = x_0$, $0 < x_0 < 1$. Here x_0 is uniquely determined by u_l and u_r.

When $\tilde{u}_r \leqslant \bar{u}_l$ the asymptotic state consists of a backward shock wave $(u_l, \tilde{u}_r)$ and a subsonic stationary wave $(\tilde{u}_r, u_r)$.

From (4.4) and (4.4′) it follows that the subcases above do not overlap. Thus given the end states u_l and u_r, there exists a unique asymptotic state connecting them. Consequently, one expects every asymptotic state to be time-asymptotically stable. This is certainly consistent with our analysis in Secs. II and III and will be wholly justified in the last section.

We next turn to the unstable case $C(x)h'(u)>0$. In this case we have, instead of (4.4) and (4.4′),

$\tilde{\bar{u}} > \bar{\tilde{u}}$ for any $u>0$, when $C(x)h'(u)>0$, $0<x<1$, (4.5)

$\bar{u}_* > u^* > 0$, when $C(x)h'(u)>0$, $0<x<1$.

There are two cases.

2.1: $u_l \leqslant u^*$. When $u_r \geqslant 0$ the asymptotic state consists of a backward wave (u_l, u_*), a subsonic stationary wave $(u_*, 0)$, and a forward rarefaction wave $(0, u_r)$.

When $u_r < 0$, the asymptotic state consists of a backward wave $(u_l, \tilde{u}_r)$ and a stationary wave $(\tilde{u}_r, u_r)$.

Case 2.2: $u_l > u^*$. Since $u_l > u^*$ there exists a unique $u_1 > 0$ with $\tilde{u}_1 = u_l$. When $u_r \geqslant u_1$ the asymptotic state consists of the stationary wave (u_l, u_1) and a forward wave (u_1, u_r).

When $u_r \geqslant \bar{u}_1$ and $\tilde{u}_r \leqslant \bar{u}_l$ the asymptotic state consists of a transonic stationary wave with a transonic stationary shock wave at $x = x_0$, $0 \leqslant x_0 \leqslant 1$.

When $\tilde{u}_r < \bar{u}_l$ the asymptotic state consists of a backward shock wave $(u_l, \tilde{u}_r)$ followed by a stationary wave $(\tilde{u}_r, u_r)$.

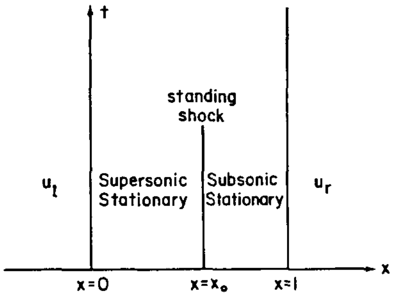

FIG. 11. $\bar{u}_* < u_l < u^*, u_r \geqslant 0$.

Unlike the stable case, we have from (4.5) and (4.5′) that the three subcases in case 2.2 overlap when $u_r \geqslant \bar{u}_1$ and $\tilde{u}_r \leqslant \bar{u}_e$ or, equivalently, $\tilde{\bar{u}} \leqslant \tilde{u}_r \leqslant \tilde{\bar{u}}_1$. In this case there exist three asymptotic states which connect given end states u_e and u_r. The one with a stationary shock wave is unstable (Sec. II). When $u^* \leqslant u_e \leqslant \bar{u}_*$ and $u_r \geqslant 0$ there are also three asymptotic states. Again, the one with a stationary shock wave is unstable.

Remark: The strong coupling hypothesis (1.3) is necessary for obtaining a finite number of asymptotic states with given end states. It is also necessary to avoid the neutral stability in the analysis in the preceding two sections. Moreover, the waves in $x<0$ $(x>1)$ with negative (positive) speed in an asymptotic state correspond to boundary layer at $x=0$ $(x=1)$ for the steady state solutions of the associated viscous equation. With (1.3), asymptotic states for (1.1) would correspond to steady state solutions of the associated viscous equation. The same would hold for general physical models. In summary, the hypothesis (1.3) is necessary for the model (1.1) to capture the nonlinear resonance phenomena observed in the nozzle flow and the MHD with a moving magnetic force.

V. EXISTENCE OF SOLUTIONS

A. Numerical scheme

In this subsection we introduce a numerical scheme for the initial value problem (1.1)

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = C(x)h(u), \quad u(x,0) = u_0(x). \qquad (5.1)$$

The scheme is based on the random choice method,[4,9] and uses elementary waves for the conservation law (1.6) and stationary waves, (1.7), as building blocks. The scheme differs from a previous one[5] in that it can treat that transonic flow. Choose an equidistributed sequence $\{a_0, a_1, a_2, ...\}$ in $(0,1)$ and mesh sizes Δx and Δt satisfying the usual Courant–Friedrichs–Lewy condition. At each time level $t = j\Delta t + 0$ the approximate solution $u_\Delta(x,t)$, $\Delta \equiv \Delta x$, consists of piecewise stationary waves with discontinuities at $x = i\Delta x$, $i = 0, \pm 1, \pm 2, ...$. Suppose that $u_\Delta(x,t)$ has been defined for $t \leqslant j\Delta t$. Then $u_\Delta(x,t)$, $j\Delta t < t \leqslant (j+1)\Delta t$ is constructed as follows: Resolve the discontinuity of $u_\Delta(x, j\Delta t)$ at $x = i\Delta x$ for the conservation law (1.6) and denote its solution as $u_{ij}((x - i\Delta x)/(t - j\Delta t))$. Here u_{ij} consists of either a shock wave or a rarefaction wave depending on $u_\Delta(i\Delta x - 0, j\Delta t) > u_\Delta(i\Delta x + 0, j\Delta t)$ or $u_\Delta(i\Delta x - 0, j\Delta t) \leqslant u_\Delta(i\Delta x + 0, j\Delta t)$. When $0 < a_{j+1} \leqslant \frac{1}{2}$ we set

$$u_\Delta(i\Delta x + 0, (j+1)\Delta t + 0) = u_{ij}(a_{j+1}\Delta x/\Delta t),$$
$$0 < a_{j+1} \leqslant \tfrac{1}{2}, \qquad (5.2a)$$

and define $u(x, (j+1)\Delta t + 0)$, $i\Delta x < x < (i+1)\Delta x$, to be the stationary wave with prescribed value (5.2a). When $\frac{1}{2} < a_{j+1} < 1$, we set instead

$$u_\Delta((i+1)\Delta x - 0, (j+1)\Delta t + 0)$$
$$= u_{(i+1)j}((1 - a_{j+1})\Delta x/\Delta t), \quad \tfrac{1}{2} < a_{j+1} < 1. \qquad (5.2b)$$

There exists the problem of not being able to solve the stationary equation (1.7) over $i\Delta x < x < (i+1)\Delta x$ with the given value (5.2). This can happen because (1.7) is singular at sonic, $u = 0$. The problem is resolved by employing the analysis in the preceding sections.

Because of the source term $C(x)h(u)$, the shock waves and rarefactions waves issued at $x = i\Delta x$, $i = 0, \pm 1, \pm 2, ...$ change their speed and strength. The above procedure approximates the shock speed and characteristic speed by constants. A wave with positive speed is accelerated to the right by Δx at the next time level if $a_{j+1}\Delta x/\Delta t$ is less than its speed, and so on. As we have seen in Secs. II and III, this may not always be a good idea. A shock wave close to being stationary should be viewed as being stationary. This will prevent unnecessary oscillation of the location of the shock wave in the stable case $C(x)h'(u) > 0$. In the unstable case the mechanism prevents large time-asymptotic error of transonic shock waves. Thus if the elementary wave issued from $(i\Delta x, j\Delta t)$ is a transonic shock wave with speed σ, $|\sigma| \leqslant C\sqrt{\Delta x}$, then we set u_{ij} to consist of a single discontinuity with zero speed. This also eliminates the problem of not being able to solve (1.7) with given data (5.2) provided that C is chosen sufficiently large, cf. (3.8). From Sec. III we know that when a rarefaction wave propagates toward the sonic state it will reflect as a compression wave. Thus when $C(x)h(u) < 0$ and a supersonic rarefaction wave is issued at $(i\Delta x, j\Delta t)$, then $u(x, (j+1)\Delta t)$, $i\Delta x < x < (i+1)\Delta x$ is defined as follows: Let u_0^i be the supersonic state which can be connected to sonic state by a stationary wave $u_0(x)$ over $i\Delta x < x < (i+1)\Delta x$. If the rarefaction wave takes values over $(0, u_0^i)$ and the random choice picks up one of these values, i.e., $0 < a_i\Delta x/\Delta t < f'(u_0^i)$, then we set, instead of (5.2),

$$u(x, (j+1)\Delta t + 0) = u_0(x), \quad i\Delta x < x < (i+1)\Delta x,$$
$$\text{when } 0 < a_i\Delta x/\Delta t < f'(u_0^i). \qquad (5.3)$$

The case of a subsonic rarefaction wave and $C(x)h(u) > 0$ is treated analogously. This completes the description of the scheme.

B. Convergence

We want to establish the convergence of the scheme for initial data having bounded total variation

$$\text{TV} \equiv \text{total variation } u_0(x) < \infty. \qquad (5.4)$$

This situation is complicated by the fact that the approximate solutions may not have uniformly bounded total variation because the stationary equation (1.7) is singular at sonic. Consider, for instance, $u_0(x) \equiv 0$ and $C(x)h(u) < 0$, $0 \leqslant x \leqslant 1$. Then the approximate solution at $t = 0$ consists of subsonic stationary waves with zero right end states. Since (1.7) is singular at sonic, the left end states of each stationary wave in $0 \leqslant x \leqslant 1$ is of the order $-(\Delta x)^{1/2}$. There are $(\Delta x)^{-1}$ such waves and their total variation is of the order

$$(\Delta x)^{1/2}(\Delta x)^{-1} = (\Delta x)^{-1/2},$$

which becomes unbounded as the mesh size $\Delta x \to 0$. Thus the analysis for hyperbolic conservation laws[6,10] needs to be refined and generalized. The convergence of the scheme is established by considering the variation of a new variable,

$$\psi(u) \equiv \int_0^u \frac{f'(v)}{(h(v))} dv \quad \text{for } u \geqslant 0,$$
$$\equiv -\int_0^u \frac{f'(v)}{(h(v))} dv \quad \text{for } u < 0. \qquad (5.5)$$

The reason for introducing $\psi(u)$ is that the stationary equation (1.7) becomes

$$\psi(u)_x = C(x) \qquad (5.6)$$

and a stationary wave $u(x)$ satisfies

$$\psi(u(x)) + \int^x C(y)dy = \text{const.} \qquad (5.6')$$

From (1.4) and (1.5), $\psi(u)$ is monotonically increasing in u. However, $\psi'(0) = 0$ and so the inverse of ψ has a singularity at $u = 0$. Though the total variation for u may not be uniformly bounded, we show that the total variation for $\psi(u)$ is uniformly bounded. We establish this for the case $C(x)h(u) < 0$ for $0 < x < 1$, the general case is shown by similar arguments. We set

$$F(t) \equiv \sum\{|\alpha|: \alpha \text{ the strength of shock waves, rarefaction waves, and subsonic stationary wave in } u_\Delta(x, j\Delta t + 0)\}$$
$$+ 3\sum\{|\alpha|: \alpha \text{ the strength of supersonic stationary waves in } u_\Delta(x, j\Delta t + 0)\}, \quad j\Delta t \leqslant t < (j+1)\Delta t. \qquad (5.7)$$

Here the strength of a wave is measured by the jump of ψ across it. From (5.6′) we have

$$F(0) \leqslant \text{total variation } \psi(u_0(x)) + \int_0^1 |C(y)|dy. \qquad (5.8)$$

We claim that

$$F(t_2) \leqslant f(t_1) \quad \text{for } t_2 \geqslant t_1 \geqslant 0. \qquad (5.9)$$

This is shown by checking each case of wave interactions.

From (5.6′) the propagation of a nontransonic wave through a stationary wave conserves F. When a transonic shock wave propagates to the left through a stationary wave the states on each side of it move away from the sonic state since $C(x)h(u) < 0$. Consequently, the strength of the shock wave increases by twice the amount of the increase in the strength of the subsonic stationary wave to the right of it. This is offset by the decrease in the strength of the supersonic stationary wave to the left of the shock wave. Notice that in (5.7) we have weighted thrice as heavily the strength of supersonic stationary waves as other waves. Thus F remains constant in this case. When a rarefaction wave is reflected as a shock wave, the stationary wave changes from being supersonic to subsonic and again F remains constant. Thus F decreases only when shock waves and rarefaction waves cancel. This proves (5.9).

From (5.8) and (5.9)

$$F(t) \leqslant \text{total variation } \psi(u_0(x)) + \int_0^1 |C(y)|dy$$
$$= O(1)\text{TV} + \int_0^1 |C(y)|dy, \quad t > 0. \qquad (5.10)$$

This shows that the total variation in X of $\psi(u_\Delta)$ is uniformly bounded. By the diagonal process there exists a sequence $\epsilon_k \to 0$ such that as $\Delta \equiv \Delta x = \epsilon_k \to 0$, $\psi(u_\Delta(x,t))$ converges locally for all rational t and all x not in the countable set to a limit function $\psi^*(x,t)$. For any given irrational t_0 we want to show that the same sequence also converges for all x not in a countable set. Since $\psi(u_\Delta(\cdot,t))$ has bounded total variation, there exists a subsequence ϵ_{kj} of ϵ_k such that $\psi(u_\Delta(x,t_0))$ converges to $\psi^*(x,t_0)$ as $\Delta = \epsilon_{kj} \to 0$ for all x not in a countable set. Let x_0 be any continuity point of $\psi^*(x,t_0)$, then the total variation of of $\psi^*(x,t_0)$ is small around $x = x_0$. The same is true of $\psi(u_\Delta(x,t_0))$, $\Delta = \epsilon_{kj}$ small enough, provided that there is no strong cancellation of shock waves and rarefactions around (x_0,t_0). Since the total amount of waves is finite, (5.10), such strong cancellation can occur only around countable points. Excluding these points, $\psi(u_\Delta(x,t))$ has small variation around (x_0,t_0) and for $\Delta = \epsilon_{kj}$ small. Since $\psi(u_\Delta(x,t))$ converges for $\Delta = \epsilon_k$ and t rational, we conclude, by choosing t arbitrarily close to t_0 that $\psi(u_\Delta(x,t))$ has small variation around (x_0,t_0) for $\Delta = \epsilon_k$ small. Consequently $\psi(u_\Delta(x,t))$ converges locally in (x,t) at (x_0,t_0). In other words $\psi(u_\Delta(x,t))$ converges to a limiting function $\psi^*(x,t)$ as $\Delta = \epsilon_k \to 0$. Since $\psi(u)$ is a monotonically increasing function of u, $u_\Delta(x,t)$ converges to a limit function $u(x,t) \equiv \psi^{-1}(\psi^*(x,t))$ as $\Delta = \epsilon_k \to 0$. After the proof of the consistency of the scheme in the next subsection, we conclude that $u(x,t)$ is an admissible weak solution of (1.1). It is well known that admissible weak solutions for the initial value problem of (1.1) are unique. This implies that $u_\Delta(x,t)$ converges to $u(x,t)$ as $\Delta \to 0$. To be able to invert ψ we need to assume that either $\psi(u)$ is unbounded ,

$$|\psi(u)| \to 0 \quad \text{as } |u| \to \infty \qquad (5.11)$$

or the initial data is not too large,

$$F(0) + \min\{\psi(u_0(-\infty)), \psi(u_0(+\infty))\}$$
$$< \min\{|\psi(-\infty)|, |\psi(+\infty)|\}. \qquad (5.12)$$

The above convergence arguments for $\psi(u_\Delta)$ based on (5.10) are similar to those of Secs. 10 and 12 of Ref. 11, where the details may be found.

C. Consistency

It remains to show that the limiting function $u(x,t)$ just obtained is a weak solution of (1.1) and (5.1), that is, the scheme is consistent. We use the technique of wave tracing.[9] Since the total variation of u_Δ may not be uniformly bounded, a certain refinement of the technique is necessary. In particular, we find it is convenient for our analysis to assume that the sequence $\{a_0,a_1,a_2,\ldots\}$ be sufficiently well equidistributed. We need to show that

$$E_\Delta \equiv \int_0^\infty \int_{-\infty}^\infty \left[u_\Delta \frac{\partial \varphi}{\partial t} + f(u_\Delta) \frac{\partial \varphi}{\partial x} - C(x)h(u_\Delta)\varphi \right]$$
$$\times (x,t)dx\,dt + \int_{-\infty}^\infty (u_\Delta \varphi)(x,0)dx \to 0 \quad \text{as } \Delta \to 0, \qquad (5.13)$$

for any smooth function $\varphi(x,t)$ with compact support in $t \geqslant 0$. The value of $u_\Delta(x,t)$, defined in Sec. V A for $t = j\Delta t$, $j = 0,1,2,\ldots$, has a natural generalization to all $t \geqslant 0$. Instead of using the exact solutions constructed in Secs. II and III, we approximate the shock locations and characteristic waves for rarefaction waves by straight lines in each time layer $(j\Delta t,(j+1)\Delta t)$. In the case of the reflection of a supersonic (subsonic) rarefaction wave into a compression wave, we set the value of u_Δ in the block by the subsonic (supersonic) stationary wave connecting the sonic state (5.3). With this, the error term E_Δ consists of two parts:

$$E_\Delta \equiv E_1 + E_2,$$

$$E_1 \equiv \sum_{j=0}^{\infty} \int_{-\infty}^{\infty} [u(x,j\Delta t) - u(x,j\Delta t+0)]\varphi(x,j\Delta t)dx.$$

Here E_2 is the error for $t \neq j\Delta t, j=0,1,2,...,$ due to the approximation just made.

We first show that $E_1 \to 0$ as $\Delta x \to 0$. Suppose that the test function φ is supported in $0 \leqslant t \leqslant T$. Partition each shock wave and rarefaction wave into subwaves so that each subwave stays together and has a constant strength in $\psi(u)$ until it is cancelled, if ever, cf. Ref. 6. As a subwave propagates according to (5.2) and (5.3), it contributes to E_1. Fix a small positive constant ϵ. In the region where a subwave takes value away from sonic by ϵ, its strength in u is dominated by its strength in $\psi(u)$ times $\epsilon^{-1/2}$. In this case, we may apply the arguments in Refs. 3 and 6 to show that the error due to the subwave is dominated by the strength in $\psi(u)$ times $(\epsilon^{-1/2})\Delta x$. This analysis fails as $\epsilon \to 0$. The refinement is explained below for each type of wave propagation.

Consider the propagation of a rarefaction subwave toward the sonic state at $x=0$. Let δ be its strength in $\psi(u)$ and $x = i\Delta x$ its location. The strength in u is of the order

$$\alpha_i \equiv (i\Delta x + \delta)^{1/2} - (i\Delta x)^{1/2}$$
$$= \delta[(i\Delta x+\delta)^{1/2} + (i\Delta x)^{1/2}]^{-1},$$

and its speed is of the order, cf. (3.7) and (3.8),

$$\lambda_i \equiv (i\Delta x)^{1/2}.$$

Let ϵ be a small fixed positive number

$$\epsilon \equiv (I\Delta x)^{1/2} \equiv (2^K \Delta x)^{1/2}.$$

The strength in u of the rarefaction wave at $i\Delta x$, $|i| > I$, is

$$O(1)\alpha_i = O(1)\delta\epsilon^{-1} = O(1)\delta.$$

Thus as we mentioned earlier, we need only to study the error due to the wave when it is located at $i\Delta x$, $|i| < I$. When it moves from $(i+1)\Delta x$ to $i\Delta x$ toward the sonic state it creates an error

$$O(1)(\Delta x - \lambda_i \Delta t)\alpha_i.$$

As it moves, without stopping, from $-I\Delta x$ to 0 the total error is

$$O(1)\sum_{i=0}^{I} (\Delta x - \lambda_i\Delta t)\alpha_i$$
$$= O(1)\sum_{i=0}^{I} \Delta x \delta (i\Delta x)^{-1/2}$$
$$= O(1)\delta(\Delta x)^{1/2} I^{1/2} = O(1)\delta\epsilon,$$

which is linear in δ and arbitrarily small as $\epsilon \to 0$. However, the wave may stay at each location $i\Delta x$ for many time steps and creates additional error. Let b_i be the time steps it stays at $i\Delta x$, then this error is

$$O(1)\sum_{i=0}^{I} b_i\alpha_i\lambda_i\Delta t = O(1)\delta b\Delta t, \quad b \equiv \sum_{i=0}^{I} b_i.$$

Thus we need to estimate $b\Delta t$, the total amount of time the wave stays between $x = I\Delta x$ and $x=0$. For the exact solution, (3.7) and (3.8), $b\Delta t$ is $O(1)(I\Delta x)^{1/2} = O(1)\epsilon$, which is again arbitrarily small. We now show that this remains so for the approximate solutions provided that the sequence $\{a_0,a_1,a_2,...\}$ is well equidistributed. We say that the sequence is equidistributed in (0,1) if $a_i \in (0,1)$, $i=0,1,2,...,$ and

$$A(L,N)/N = |L| + O(N)/N, \tag{5.14}$$

as $N \to \infty$. Here $A(L,N)$ denotes the number of n, $0 \leqslant n \leqslant N$, a_n belongs to any subintegral L of (0.1), and $|L|$ the length of L. The best equidistributed sequence has the estimate

$$O(N) = O(1)\log N,$$

and the average is

$$O(N) = O(1)N^{1/2}.$$

For our purpose we require that

$$O(N) = O(1)N^\alpha, \quad \text{for some } 0<\alpha<\tfrac{2}{3}. \tag{5.14'}$$

This includes most of the equidistributed sequences. We now continue the error analysis for rarefaction waves. Let C_k be the number of time steps the wave spent between $x = 2^k\Delta x$ and $x = 2^{k+1}\Delta x$,

$$C_k \equiv \sum_{i=2^k}^{2^{k+1}} b_i.$$

The wave speed lies in $((2^k\Delta x)^{1/2}, (2^{k+1}\Delta x)^{1/2})$. It takes $2^{k+1} - 2^k = 2^k$ jumps for the wave to move from $x = 2^{k+1}\Delta x$ to $x = 2^k\Delta x$. The wave jumps at time $j\Delta t$ if $a_j\Delta x$ is less than $\lambda_j\Delta t$. Since $\lambda_i \in ((2^k\Delta x)^{1/2}, (2^{k+1}\Delta x)^{1/2})$ we have from (5.14) and (5.14') that

$$(2^k\Delta x)^{1/2}\frac{\Delta t}{\Delta x} + O(1)\frac{(C_k)^\alpha}{C_k}$$
$$\leqslant \frac{2^k}{C_k} \leqslant (2^{k+1}\Delta x)^{1/2}\frac{\Delta t}{\Delta x} + O(1)\frac{(C_k)^\alpha}{C_k}.$$

We apply this to k sufficiently large

$$O(1)(C_k)^\alpha \leqslant 2^k \tag{5.15}$$

so that the above yields

$$C_k = O(1)(2^k/\Delta x)^{1/2}.$$

This in turn implies that (5.15) is equivalent to

$$2^k \geqslant O(1)(\Delta x) - \alpha/(2-\alpha). \tag{5.15'}$$

For those k satisfying (5.15') the error is

$$O(1)\delta\Delta t\sum_{k=0}^{K} C_k = O(1)\delta\Delta t(\Delta x)^{-1/2}\sum_{k=0}^{K}(2^k)^{1/2}$$
$$= O(1)\delta\Delta t(\Delta x)^{-1/2}(2^K)^{1/2}$$
$$= O(1)\delta(I\Delta x)^{1/2} = O(1)\delta\epsilon,$$

which again is linear in δ and becomes small as $\epsilon \to 0$. For those k which fail to satisfy (5.15') we derive a cruder estimate as follows: When a_j lies in $(0,\Delta x)$, the wave definitely moves toward the sonic. It takes 2^k moves to reach sonic from $x = 2^k\Delta x$. Let $\bar{k}$ be the smallest integer satisfying

(5.15′) and $\bar{C}$ the number of time steps the wave spends between $x = 2^{\bar{k}}\Delta x$ and $x = 0$. We have from (5.14) and (5.14′) that

$$2^{\bar{k}}/\bar{C} \geqslant (\Delta x)^{1/2} + O(1)[(\bar{C})^{\alpha}/\bar{C}]$$

provided that

$$2^{\bar{k}} \geqslant O(1)(\bar{C})^{\alpha}. \tag{5.16}$$

It follows that

$$\bar{C} \leqslant 2^{\bar{k}}(\Delta x)^{-1/2}$$

and (5.16) is equivalent to

$$2^{\bar{k}} \geqslant O(1)(\Delta x)^{-\alpha/(2-2\alpha)}. \tag{5.16′}$$

The error is dominated by

$$O(1)\delta\bar{C}\Delta t = O(1)\delta 2^{\bar{k}}(\Delta x)^{1/2}.$$

We want the above two cases to be exhaustive, in other words, the smallest integer $k = \bar{k}$ satisfying (5.15′) also satisfies (5.16′). This is so because

$$(\Delta x)^{-\alpha/(2-2\alpha)} > (\Delta x)^{-\alpha/(2-\alpha)}$$

for $\Delta x < 1$ and $\alpha > 0$. The last error is then

$$O(1)\delta 2^{\bar{k}}(\Delta x)^{1/2} = O(1)\delta(\Delta x)^{1/2-\alpha/(2-\alpha)}$$
$$= O(1)\delta(\Delta x)^{(2-3\alpha)/(2-\alpha)},$$

which is linear in δ and tends to zero as $\Delta x \to 0$ because $0 < \alpha < \frac{2}{3}$, (5.14′). Our analysis shows that the total error due to the rarefaction wave is linear in δ and tends to zero as $\Delta x \to 0$.

Consider next the propagation of a nontransonic shock wave. If the shock wave does not interact with other shock or rarefaction waves then we may use the above argument for rarefaction wave for the error analysis. The change in shock speed due to interaction has to be estimated in the general analysis. Since wave strength in u may not be uniformly bounded near sonic, the total amount of waves interacting with the shock wave may be large. However, only the oscillation of the shock speed is relevant for error analysis and it depends on the oscillation of waves interacting with the shock wave. Away from the sonic state the oscillation of waves is dominated by their variation in u, which is equivalent to the total variation in $\psi(u)$ and is thereby uniformly bounded. Near the sonic state, the oscillation is small by definition. In short, the change in shock speed can be controlled and we obtain similar error bound, which is linear in the wave strength of the shock wave in $\psi(u)$ and tends to zero as $\Delta x \to 0$.

Finally we consider a transonic shock wave, which may oscillate around the sonic state and so the above arguments do not apply. However, we will show that the total amount of transonic shock waves in both u and $\psi(u)$ is nearly uniformly bounded. Precisely, we have the following: Suppose that $C(x)h(u) > 0$, $0 \leqslant x \leqslant 1$. We may artificially push all shock waves and rarefaction waves to $x = 0$ without increasing the functional $F(t)$. This moves the end states of waves away from sonic. Given any fixed $\epsilon > 0$, transonic shock waves in u increases by at least $C\epsilon^{1/2}$, C some positive constant. We have a transonic shock wave that has strength greater than $C\epsilon^{1/2}$, its strength measured in u is $O(1)\delta\epsilon^{-1/2}$, δ its strength in $\psi(u)$. We thus conclude that the total strength of transonic shock waves in $x > \epsilon$ for any given time is

$$O(1)F(t)\epsilon^{-1/2} = O(1)F(0)\epsilon^{-1/2}.$$

Thus we may use previous arguments and a standard one,[3,6] to obtain the desired error analysis for transonic shock waves in $x > \epsilon$. The error analysis for the region $0 \leqslant x \leqslant \epsilon$ is trivial because the area $\{(x,t): 0 \leqslant x \leqslant \epsilon, 0 \leqslant t \leqslant T\}$ goes to zero as $\epsilon \to 0$ and because u_Δ are uniformly bounded.

We consider next the error term E_2. Only the error due to the reflection of rarefaction waves to compression waves (5.3) is analyzed; the rest follows from arguments similar to those above. Note that rarefaction waves may disappear due to cancellation but not be created. Different rarefaction waves at $t = 0$ reach sonic at different x later. Consequently, the total area of the region of wave reflection, where (5.3) applies, in the (x,t) plane is no larger than Δt and the corresponding error in E_2 is $O(1)\Delta t$.

Summing up the above error analysis we conclude that E_Δ of (5.13) satisfies

$$E_\Delta = O(1)(\epsilon + O(\Delta x)) + O(1)F(0)O(\Delta x)\epsilon^{-1/2},$$

which is $O(1)\epsilon$ as $\Delta x \to 0$. Since ϵ is arbitrarily chosen, we have $E_\Delta \to 0$ as $\Delta x \to 0$. Thus the limiting function obtained in Sec. V B is a weak solution.

The solution obtained by the scheme has the property that $\psi(u(x,t))$ has bounded total variation in x for $t \geqslant 0$, (5.10), and that $u(x,t)$ is uniformly bounded. It is well known that the nonlinearity $f''(u) \neq 0$ has a regularizing effect in that if $u(x,t)$ is bounded and measurable then $u(x,t')$ is of bounded local variation for $t' > t$, $|t' - t|$ small. Thus our solution $u(x,t)$ is of locally bounded variation in x for all t. In fact it is of bounded variation because $u(x,0)$ is of bounded variation and (1.1) is a conservation law for $x \in [0,1]$. We summarize our results in the following theorem.

Theorem 5.1: Suppose that the initial data $u(x,0)$ have bounded total variation (TV) and (5.11) or (5.12) holds. Then (1.1) has a unique global solution $u(x,t)$ which is of bounded variation in x for all t and is the limit of approximate solutions $u_\Delta(x,t)$ as $\Delta \to 0$.

The uniqueness of the solution follows from general theory of hyperbolic conservation law. As we mentioned above, due to the nonlinearity $f''(u) \neq 0$, we may relax somewhat the hypothesis that $u(x,0)$ is of bounded total variation. This is, however, not of importance in the present context. Our primary concern is the qualitative behavior of nonlinear waves and efficient numerical computation.

VI. QUALITATIVE BEHAVIOR OF SOLUTIONS

In this section we study the regularity and large-time behavior of solutions of (1.1). Since solutions $u(x,t)$ obtained in the last section are of bounded variation, the regularity theory for hyperbolic conservation laws (see Ref. 3 and references therein) can be straightforwardly generalized and applied here. There are countable points of wave interactions, a countable set of Lipschitz continuous curves of shock wave, and the solution is continuous elsewhere. As in Sec. V B, we may refine the argument of Secs. 12–15 of Ref. 3

to show that our scheme in Sec. V A converges locally for points of continuity of the solution and computes shock waves sharply.

We next turn to the large-time behavior and show that the solution tends to an asymptotic state constructed in Sec. IV. Since the solution $u(x,t)$ is of bounded variation and at $x = \pm\infty$ (1,1) is the conservation law (1.6), the end states

$$u_e \equiv u(-\infty,t), \quad u_r \equiv u(+\infty,t)$$

are time invariants. From the analysis in Sec. III and the first part of Sec. V C, we see that there is no rarefaction in $0 \leqslant x \leqslant 1$ after finite time. Similarly nontransonic shock waves also move out of $0 \leqslant x \leqslant 1$ in finite time. It is obvious, that, without the presence of rarefaction waves and nontransonic shock waves in $0 \leqslant x \leqslant 1$, there exists at most one transonic shock wave in $0 \leqslant x \leqslant 1$. This is so under the present simplified assumption of $C(x) \neq 0$ for $0 \leqslant x \leqslant 1$, cf. Ref. 2. In the region of $x \notin [0,1]$, (1.1) is the same as the conservation law (1.6), for which it is well known that waves eventually become noninteracting.[4] In the present situation, only one shock or rarefaction wave with negative (positive) speed eventually survives in $x < 0$ $(x > 1)$. In summary, $u(x,t)$ tends to a noninteracting wave pattern, an asymptotic state of Sec. IV as $t \to \infty$. The rate of this convergence is $t^{-1/2}$ as in the case of conservation law.

[1]T.-P. Liu, "Non-linear stability and instability of transonic gas flow through a nozzle," Commun. Math. Phys. **83**, 243 (1982).

[2]C.-Z. Li and T.-P. Liu, "Asymptotic states for hyperbolic conservation laws with a moving source," Adv. Appl. Math. **4**, 353 (1983).

[3]T.-P. Liu, "Admissible solutions of hyperbolic conservation laws," Mem. Am. Math. Soc. **240**, 1 (1981).

[4]J. Glimm, "Solutions in the large for non-linear hyperbolic systems of equations," Commun. Pure Appl. Math. **18**, 697 (1965).

[5]J. Glimm, G. Marshall, and B. Plohr, "A generalized Riemann problem for quasi-one-dimensional gas flows," Adv. Appl. Math. **5**, 1 (1984).

[6]T.-P. Liu, "The deterministic version of the Glimm scheme," Commun. Math. Phys. **57**, 135 (1977).

[7]T.-P. Liu, "Resonance for quasilinear hyperbolic equation," Bull. Am. Math. Soc. **6**, 463 (1982).

[8]H. Glaz and T.-P. Liu, "The asymptotic analysis of wave interactions and numerical calculation of transonic nozzle flow," Adv. Appl. Math. **5**, 111 (1984).

[9]L. Huang, and T.-P. Liu, "A conservative, piecewise-steady difference scheme for transonic nozzle flow," Comput. Math. Appl. A **12**, 377 (1986).

[10]T.-P. Liu, "Quasilinear hyperbolic systems," Commun. Math. Phys. **68**, 141 (1979).

[11]S. K. Fok, "Extension of Glimm's method to the problem of gas flow in a duct of variable cross-section," Lawrence Berkeley Laboratory Report LBL-12322, 1980.

[12]P. Embid, J. Goodman, and J. Majda, "Multiple steady states for 1-D transonic flow," SIAM J. Sci. State Comp. **5**, 1, 21 (1984).

[13]G. Kreiss and H.-O. Kreiss, "Convergence to steady state of solutions of Burgers' equation," I CASE Report No. 86-18, 1985, pp. 1–17.

[14]B. van Leer, "On the relation between the upwind-differencing schemes of Godunov, Enquist–Osher, and Roe," I CASE Report No. 81-11, 1981.

[15]B. Whitham, *Linear and Non-linear Waves* (Wiley, New York, 1974).

Reflection of waves in nonlinear integrable systems

V. K. Mel'nikov
Joint Institute for Nuclear Research, Dubna, USSR

(Received 17 April 1987; accepted for publication 10 June 1987)

Solutions of the boomeron type are found in two nonlinear integrable systems describing the interaction of a long wave with a short wave packet. These solutions follow from two-soliton solutions if certain additional conditions are imposed on their parameters. The results are relevant to some problems of plasma physics, solid-state physics, hydrodynamics, etc.

I. INTRODUCTION

The present paper is devoted to the following phenomenon. In the nonlinear integrable systems considered below we find solutions having asymptotics of the one-soliton type as $t \to \pm\infty$. However, the sets of essential parameters of these solitons are different. In particular, these sets of parameters may be chosen so that the solitons mentioned above would have opposite directions of motion. In this case, each solution of this kind describes a wave going from infinity. Then, the parameters of this wave change (in particular, the direction of propagation). As a result, the wave begins propagating in the opposite direction, and finally, goes back to where it started. In other words, it's as though the wave is reflected.

The following fact is to be noted. Let us take an arbitrary small $\epsilon > 0$. Let t_- be such that at $t < t_-$ the difference between our solution and the first soliton is less than ϵ. Then let t_+ be such that at $t > t_+$ the difference between the solution considered and the second soliton is also less than ϵ. It turns out that t_- and t_+ may be chosen so that the difference $t_+ - t_-$ as $\epsilon \to 0$ would be of an order of $-\ln \epsilon$. This means that the rearrangement process of one soliton into another takes place during a comparatively small interval of time, and consequently, the calculation of such solutions by computer is a very nontrivial problem.

The aforesaid will be demonstrated by two examples of nonlinear evolution systems, one describing the interaction of waves on the xy plane; and the second, the interaction of waves on the x axis.

II. INTERACTION OF TWO WAVES ON THE xy PLANE

We proceed with the following system of equations:

$$3\frac{\partial^2 u}{\partial y^2} - \frac{\partial}{\partial x}\left[\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(3u^2 + \frac{\partial^2 u}{\partial x^2} + 8\varkappa|\varphi|^2\right)\right] = 0,$$
$$\mathrm{i}\frac{\partial \varphi}{\partial y} = u\varphi + \frac{\partial^2 \varphi}{\partial x^2}, \tag{2.1}$$

where $\mathrm{i} = \sqrt{-1}$, describing (under certain conditions) the interaction of a long wave with a short wave packet propagating on the xy plane at an angle to each other. Here u is the long wave amplitude, φ is the complex short wave envelope, and the parameter $\varkappa$ satisfies the condition $\varkappa^2 = 1$. One can easily be convinced that system (2.1) has solutions of the form

$$u = \frac{2\mu^2}{\cosh^2[\mu(x + 2\nu y - \tau t)]},$$
$$\varphi = a\frac{\exp[\mathrm{i}\nu(x + 2\nu y) + \mathrm{i}\sigma t]}{\cosh[\mu(x + 2\nu y - \tau t)]}\exp[-\mathrm{i}(\mu^2 + \nu^2)y], \tag{2.2}$$

where the real parameters μ, ν, and τ and the complex quantity a satisfy the sole condition

$$[\tau - 4(\mu^2 - 3\nu^2)]\mu^2 = 4\varkappa|a|^2, \tag{2.3}$$

and consequently, these solutions may exist only if the condition $[\tau - 4(\mu^2 - 3\nu^2)]\varkappa \geqslant 0$ is fulfilled. In this case the parameter σ may acquire arbitrary real values. The inverse scattering method[1,2] allows one to consider the interaction of an arbitrary number of waves of the form (2.2).

For our aim it is sufficient to consider the interaction of two such waves.

We take the functions D and Φ of the form

$$\begin{aligned} D = {} & 1 + \alpha_1 \exp[2\mu_1(x + 2\nu_1 y - \tau_1 t)] + \alpha_2 \exp[2\mu_2(x + 2\nu_2 y - \tau_2 t)] + \gamma_0 \exp[2\mu_1(x + 2\nu_1 y - \tau_1 t) \\ & + 2\mu_2(x + 2\nu_2 y - \tau_2 t)] + 2\delta_0 \exp[\mu_1(x + 2\nu_1 y - \tau_1 t) + \mu_2(x + 2\nu_2 y - \tau_2 t)]\cos\theta, \\ \Phi = {} & 2a_1\{1 + \beta_2 \exp[2\mu_2(x + 2\nu_2 y - \tau_2 t)]\}\exp[\mu_1(x + 2\nu_1 y - \tau_1 t)]\exp[\mathrm{i}\nu_1(x + 2\nu_1 y) + \mathrm{i}\sigma_1 t - \mathrm{i}(\mu_1^2 + \nu_1^2)y] \\ & + 2a_2\{1 + \beta_1 \exp[2\mu_1(x + 2\nu_1 y - \tau_1 t)]\}\exp[\mu_2(x + 2\nu_2 y - \tau_2 t)] \\ & \times \exp[\mathrm{i}\nu_2(x + 2\nu_2 y) + \mathrm{i}\sigma_2 t - \mathrm{i}(\mu_2^2 + \nu_2^2)y], \end{aligned} \tag{2.4}$$

where

$$\alpha_1 = \frac{\varkappa|a_1|^2}{(\mu_3^2 - 3\nu_3^2)\mu_1\mu_3}, \quad \alpha_2 = \frac{\varkappa|a_2|^2}{(\mu_4^2 - 3\nu_4^2)\mu_2\mu_4},$$

$$\sigma_1 = -4(3\mu_1^2 - \nu_1^2)\nu_1 + 4(3\mu_3^2 - \nu_3^2)\nu_3,$$

$$\sigma_2 = -4(3\mu_2^2 - \nu_2^2)\nu_2 + 4(3\mu_4^2 - \nu_4^2)\nu_4,$$

 0022-2488/87/112603-07$02.50

$$\tau_1 = 4(\mu_1^2 - 3\nu_1^2) + 4(\mu_3^2 - 3\nu_3^2)\mu_1^{-1}\mu_3,$$
$$\tau_2 = 4(\mu_2^2 - 3\nu_2^2) + 4(\mu_4^2 - 3\nu_4^2)\mu_2^{-1}\mu_4,$$
$$\theta = (\nu_2 - \nu_1)x + (\mu_1^2 - \nu_1^2 - \mu_2^2 + \nu_2^2)y + (\sigma_2 - \sigma_1)t + \theta_0, \tag{2.5}$$
$$\omega_1 = \mu_1 + i\nu_1, \quad \omega_2 = \mu_2 + i\nu_2,$$
$$\omega_3 = \mu_3 + i\nu_3, \quad \omega_4 = \mu_4 + i\nu_4,$$
$$\beta_1 = \alpha_1 \frac{\omega_2 - \omega_1}{\omega_2 + \overline{\omega}_1} \frac{\overline{\omega}_4^3 - \overline{\omega}_3^3}{\overline{\omega}_4^3 + \omega_3^3}, \quad \beta_2 = \alpha_2 \frac{\omega_1 - \omega_2}{\omega_1 + \overline{\omega}_2} \frac{\overline{\omega}_3^3 - \overline{\omega}_4^3}{\overline{\omega}_3^3 + \omega_4^3},$$
$$\gamma_0 = \alpha_1\alpha_2 \left|\frac{\omega_1 - \omega_2}{\omega_1 + \overline{\omega}_2}\right|^2 \left|\frac{\omega_3^3 - \omega_4^3}{\omega_3^3 + \overline{\omega}_4^3}\right|^2,$$
$$\delta_0^2 = \frac{|\omega_1 + \overline{\omega}_1|\,|\omega_2 + \overline{\omega}_2|\,|\omega_3^3 + \overline{\omega}_3^3|\,|\omega_4^3 + \overline{\omega}_4^3|}{|\omega_1 + \overline{\omega}_2|^2|\omega_3^3 + \overline{\omega}_4^3|^2}|\alpha_1\alpha_2|.$$

Here and everywhere the bar above any quantity means a complex conjugation. According to the results in Ref. 2 the functions

$$u = 2\frac{\partial^2}{\partial x^2}\ln D, \quad \varphi = \frac{\Phi}{D} \tag{2.6}$$

satisfy system (2.1), i.e., are its solutions. One can easily see that if the conditions

$$[\tau_1 - 4(\mu_1^2 - 3\nu_1^2)]\varkappa > 0, \quad [\tau_2 - 4(\mu_2^2 - 3\nu_2^2)]\varkappa > 0, \tag{2.7}$$

$$|\omega_1 + \overline{\omega}_1|\,|\omega_2 + \overline{\omega}_2|\,|\omega_3^3 + \overline{\omega}_3^3|\,|\omega_4^3 + \overline{\omega}_4^3| \leqslant |\omega_1 + \overline{\omega}_2|^2|\omega_3^3 + \overline{\omega}_4^3|^2 \tag{2.8}$$

are fulfilled, we have $\alpha_1 \geqslant 0$, $\alpha_2 \geqslant 0$, $0 \leqslant \delta_0^2 \leqslant \alpha_1\alpha_2$, $\gamma_0 \geqslant 0$ and consequently, the function D is positive at any real x, y, and t. This means that the functions u and φ defined by (2.6) have no singularities at any real values of x, y, and t.

Let us now analyze the behavior of this solution. Consider first the case $(\omega_1 - \omega_2)(\omega_3^3 - \omega_4^3) \neq 0$. Then, at $\nu_1 \neq \nu_2$ and at any fixed t in the region $\mu_2(x + 2\nu_2 y - \tau_2 t) \gg 1$ our solution has asymptotics of the form

$$u \sim u_1^+ = \frac{2\mu_1^2}{\cosh^2[\mu_1(x + 2\nu_1 y - \tau_1 t) + \delta_1^+]},$$
$$\varphi \sim \varphi_1^+ = a_1^+ \frac{\exp[i\nu_1(x + 2\nu_1 y) + i\sigma_1 t]}{\cosh[\mu_1(x + 2\nu_1 y - \tau_1 t) + \delta_1^+]} \times \exp[-i(\mu_1^2 + \nu_1^2)y], \tag{2.9}$$

where

$$\delta_1^+ = \tfrac{1}{2}(\ln\gamma_0 - \ln\alpha_2), \quad a_1^+ = a_1\alpha_2^{-1}\beta_2\exp(-\delta_1^+),$$

and in the region $\mu_2(x + 2\nu_2 y - \tau_2 t) \ll -1$ the asymptotics

$$u \sim u_1^- = \frac{2\mu_1^2}{\cosh^2[\mu_1(x + 2\nu_1 y - \tau_1 t) + \delta_1^-]},$$
$$\varphi \sim \varphi_1^- = a_1^- \frac{\exp[i\nu_1(x + 2\nu_1 y) + i\sigma_1 t]}{\cosh[\mu_1(x + 2\nu_1 y - \tau_1 t) + \delta_1^-]} \times \exp[-i(\mu_1^2 + \nu_1^2)y], \tag{2.10}$$

are fulfilled, where

$$\delta_1^- = \tfrac{1}{2}\ln\alpha_1, \quad a_1^- = a_1\exp(-\delta_1^-).$$

By virtue of (2.5) we get

$$\delta_1 = \delta_1^+ - \delta_1^- = \ln\left|\frac{\omega_1 - \omega_2}{\omega_1 + \overline{\omega}_2}\right| + \ln\left|\frac{\omega_3^3 - \omega_4^3}{\omega_3^3 + \overline{\omega}_4^3}\right|,$$
$$|a_1^+| = |a_1^-|.$$

Now moving on the xy plane to infinity along the straight line $\mu_1(x + 2\nu_1 y - \tau_1 t) + \delta_1^+ = 0$, asymptotics (2.9) are seen to hold as $y \to \infty$ if $(\nu_2 - \nu_1)\mu_2 > 0$ and vice versa if $(\nu_2 - \nu_1)\mu_2 < 0$ asymptotics (2.9) hold as $y \to -\infty$. Analogously, moving to infinity along the straight line $\mu_1(x + 2\nu_1 y - \tau_1 t) + \delta_1^- = 0$ we easily find that asymptotics (2.10) are valid at $(\nu_2 - \nu_1)\mu_2 > 0$ as $y \to -\infty$ and if $(\nu_2 - \nu_1)\mu_2 < 0$, asymptotics (2.10) are valid as $y \to \infty$.

Then one can easily be convinced that at $\nu_1 \neq \nu_2$ and at any fixed t in the region $\mu_1(x + 2\nu_1 y - \tau_1 t) \gg 1$ the solution considered has the asymptotics

$$u \sim u_2^+ = \frac{2\mu_2^2}{\cosh^2[\mu_2(x + 2\nu_2 y - \tau_2 t) + \delta_2^+]},$$
$$\varphi \sim \varphi_2^+ = a_2^+ \frac{\exp[i\nu_2(x + 2\nu_2 y) + i\sigma_2 t]}{\cosh[\mu_2(x + 2\nu_2 y - \tau_2 t) + \delta_2^+]} \times \exp[-i(\mu_2^2 + \nu_2^2)y], \tag{2.11}$$

where

$$\delta_2^+ = \tfrac{1}{2}(\ln\gamma_0 - \ln\alpha_1), \quad a_2^+ = a_2\alpha_1^{-1}\beta_1\exp(-\delta_2^+),$$

and in the region $\mu_1(x + 2\nu_1 y - \tau_1 t) \ll -1$ it has asymptotics of the form

$$u \sim u_2^- = \frac{2\mu_2^2}{\cosh^2[\mu_2(x + 2\nu_2 y - \tau_2 t) + \delta_2^-]},$$
$$\varphi \sim \varphi_2^- = a_2^- \frac{\exp[i\nu_2(x + 2\nu_2 y) + i\sigma_2 t]}{\cosh[\mu_2(x + 2\nu_2 y - \tau_2 t) + \delta_2^-]} \times \exp[-i(\mu_2^2 + \nu_2^2)y], \tag{2.12}$$

where

$$\delta_2^- = \tfrac{1}{2}\ln\alpha_2, \quad a_2^- = a_2\exp(-\delta_2^-).$$

By virtue of (2.5) we find that

$$\delta_2 = \delta_2^+ - \delta_2^- = \ln\left|\frac{\omega_1 - \omega_2}{\omega_1 + \overline{\omega}_2}\right| + \ln\left|\frac{\omega_3^3 - \omega_4^3}{\omega_3^3 + \overline{\omega}_4^3}\right|,$$
$$|a_2^+| = |a_2^-|.$$

Moving on the xy plane to infinity along the straight line $\mu_2(x + 2\nu_2 y - \tau_2 t) + \delta_2^+ = 0$, we find that asymptotics (2.11) are valid at $(\nu_1 - \nu_2)\mu_1 > 0$ as $y \to \infty$; vice versa if $(\nu_1 - \nu_2)\mu_1 < 0$ asymptotics (2.11) are fulfilled as $y \to -\infty$. Analogously, moving to infinity along the straight line $\mu_2(x + 2\nu_2 y - \tau_2 t) + \delta_2^- = 0$, we easily get that at $(\nu_1 - \nu_2)\mu_1 > 0$ as $y \to -\infty$ asymptotics (2.12) hold and if $(\nu_1 - \nu_2)\mu_1 < 0$ asymptotics (2.12) are valid as $y \to \infty$.

Thus at $(\omega_1 - \omega_2)(\omega_3^3 - \omega_4^3) \neq 0$ and $\nu_1 \neq \nu_2$ our solution describes the interaction of two solitary waves of the form (2.2) propagating on the xy plane at an angle to each other. The nonlinear nature of the interaction leads to a large distortion of these waves in the vicinity of the intersection point of straight lines

$$x + 2\nu_1 y - \tau_1 t + \tfrac{1}{2}(\delta_1^+ + \delta_1^-)\mu_1^{-1} = 0,$$
$$x + 2\nu_2 y - \tau_2 t + \tfrac{1}{2}(\delta_2^+ + \delta_2^-)\mu_2^{-1} = 0.$$

However, if one moves to infinity along the crest of any of the interacting waves, each wave acquires the form indicated above. Far away from the interaction region the result of interaction leads to a phase shift of both waves.

In the case when $(\omega_1 - \omega_2)(\omega_3^3 - \omega_4^3) \neq 0$ and $\nu_1 = \nu_2$, our solution describes the interaction of two waves only if $\tau_1 \neq \tau_2$. Both waves in this case propagate in the same direction if $\tau_1\tau_2 > 0$ and in the opposite directions if $\tau_1\tau_2 < 0$. Note that if $(\tau_1 - \tau_2)\mu_2 > 0$ as $t \to \infty$ the asymptotics of one of the waves have the form (2.9) and as $t \to -\infty$ its asymptotics have the form (2.10). In the opposite situation, if $(\tau_1 - \tau_2)\mu_2 < 0$, then as $t \to \infty$ the asymptotics of this wave have the form (2.10) and as $t \to -\infty$ they have the form (2.9). Note further that if $(\tau_2 - \tau_1)\mu_1 > 0$, then as $t \to \infty$ the asymptotics of the second of the interacting waves have the form (2.11) and as $t \to -\infty$ its asymptotics have the form (2.12). In the opposite situation, if $(t_2 - \tau_1)\mu_1 < 0$, then as $t \to \infty$ asymptotics (2.12) hold and as $t \to -\infty$ asymptotics (2.11) are valid. The distortion of these waves in this case is maximal at the time moment

$$t_0 = \frac{(\delta_1^+ + \delta_1^-)\mu_2 - (\delta_2^+ + \delta_2^-)\mu_1}{2(\tau_1 - \tau_2)\mu_1\mu_2}$$

and tends to zero as $t \to \pm\infty$.

Finally, at $(\omega_1 - \omega_2)(\omega_3^3 - \omega_4^3) \neq 0$ and $\nu_1 = \nu_2$, $\tau_1 = \tau_2$ our solution describes one solitary wave formed by two merged waves that move as one. However, the shape of this wave differs greatly from that of the initial waves. For instance, let $\omega_1 = \mu + i\nu$, $\omega_2 = 2\mu + i\nu$, and the quantities $\omega_3 = \mu_3 + i\nu_3$ and $\omega_4 = \mu_4 + i\nu_4$ be chosen so the conditions $(\mu_3^2 - 3\nu_3^2)\mu_3 = 3\mu^3$ and $(\mu_4^2 - 3\nu_4^2)\mu_4 = 0$ are fulfilled. Assuming then that $a_2 = 0$, one can easily obtain that the expressions for D and Φ are

$$\begin{aligned} D &= 1 + \alpha_1 \exp[2\mu(x + 2\nu y - \tau t)] \\ &\quad + \alpha_2 \exp[4\mu(x + 2\nu y - \tau t)] \\ &\quad + \tfrac{1}{9}\alpha_1\alpha_2 \exp[6\mu(x + 2\nu y - \tau t)], \\ \Phi &= 2a_1\{1 - \tfrac{1}{3}\alpha_2 \exp[4\mu(x + 2\nu y - \tau t)]\} \\ &\quad \times \exp[\mu(x + 2\nu y - \tau t)] \\ &\quad \times \exp[i\nu(x + 2\nu y) + i\sigma t - i(\mu^2 + \nu^2)y], \end{aligned}$$

where

$$\alpha_1 = \varkappa|a_1|^2/3\mu^4, \quad \tau = 4(4\mu^2 - 3\nu^2),$$
$$\sigma = -4(3\mu^2 - \nu^2)\nu + 4(3\mu_3^2 - \nu_3^2)\nu_3,$$

and the quantity α_2 can be chosen arbitrarily. Assuming $\alpha_2 = \frac{1}{3}\alpha_1^2$ we find that

$$D = \{1 + \tfrac{1}{3}\alpha_1 \exp[2\mu(x + 2\nu y - \tau t)]\}^3.$$

At $\varkappa = 1$ and $a_1 \neq 0$ we have $\alpha_1 > 0$, and consequently, $D > 1$ at any real values of x, y, and t. Thus, the solution under consideration has the form

$$\begin{aligned} u &= \frac{6\mu^2}{\cosh^2[\mu(x + 2\nu y - \tau t) + \delta]}, \\ \varphi &= a\frac{\sinh[\mu(x + 2\nu y - \tau t) + \delta]}{\cosh^2[\mu(x + 2\nu y - \tau t) + \delta]} \\ &\quad \times \exp[i\nu x + i\sigma t - i(\mu^2 - \nu^2)y], \end{aligned}$$

where

$$\delta = \tfrac{1}{2}\ln(\alpha_1/3) = \ln(|a_1|/3\mu^2), \quad a = a_1\exp(-\delta),$$

and consequently, the equality $|a| = 3\mu^2$ is valid.

The situation changes radically if the condition $(\omega_1 - \omega_2)(\omega_3^3 - \omega_4^3) = 0$ is fulfilled. Consider first the case $\nu_1 \neq \nu_2$. Then, at any fixed t in the region $\mu_2(x + 2\nu_2 y - \tau_2 t) \gg 1$ our solution has the zero asymptotics and in the region $\mu_2(x + 2\nu_2 y - \tau_2 t) \ll -1$ the following asymptotics hold:

$$\begin{aligned} u &\sim \frac{2\mu_1^2}{\cosh^2[\mu_1(x + 2\nu_1 y - \tau_1 t) + \delta_1]}, \\ \varphi &\sim \hat{a}_1 \frac{\exp[i\nu_1(x + 2\nu_1 y) + i\sigma_1 t]}{\cosh[\mu_1(x + 2\nu_1 y - \tau_1 t) + \delta_1]} \\ &\quad \times \exp[-i(\mu_1^2 + \nu_1^2)y], \end{aligned} \tag{2.13}$$

where $\delta_1 = \frac{1}{2}\ln\alpha_1$ and $\hat{a}_1 = a_1\exp(-\delta_1)$. Then, at any fixed t in the region $\mu_1(x + 2\nu_1 y - \tau_1 t) \gg 1$ the solution considered has the zero asymptotics and in the region $\mu_1(x + 2\nu_1 y - \tau_1 t) \ll -1$ the asymptotics

$$\begin{aligned} u &\sim \frac{2\mu_2^2}{\cosh^2[\mu_2(x + 2\nu_2 y - \tau_2 t) + \delta_2]}, \\ \varphi &\sim \hat{a}_2 \frac{\exp[i\nu_2(x + 2\nu_2 y) + i\sigma_2 t]}{\cosh[\mu_2(x + 2\nu_2 y - \tau_2 t) + \delta_2]} \\ &\quad \times \exp[-i(\mu_2^2 + \nu_2^2)y] \end{aligned} \tag{2.14}$$

are fulfilled, where $\delta_2 = \frac{1}{2}\ln\alpha_2$ and $\hat{a}_2 = a_2\exp(-\delta_2)$. According to (2.5) the equalities

$$\begin{aligned} 4\varkappa|\hat{a}_1|^2 &= 4(\mu_3^2 - 3\nu_3^2)\mu_1\mu_3 = [\tau_1 - 4(\mu_1^2 - 3\nu_1^2)]\mu_1^2, \\ 4\varkappa|\hat{a}_1|^2 &= 4(\mu_4^2 - 3\nu_4^2)\mu_2\mu_4 = [\tau_2 - 4(\mu_2^2 - 3\nu_2^2)]\mu_2^2 \end{aligned} \tag{2.15}$$

are valid, which are analogous to relations (2.3). Thus, moving on the xy plane to infinity on the straight line $\mu_1(x + 2\nu_1 y - \tau_1 t) + \delta_1 = 0$, we find that at $(\nu_2 - \nu_1)\mu_2 > 0$ and $y \to \infty$ our solution has zero asymptotics and as $y \to -\infty$ asymptotics (2.13) are valid, vice versa if $(\nu_2 - \nu_1)\mu_2 < 0$, then moving to infinity along this straight line we see that as $y \to -\infty$ our solution has the zero asymptotics and as $y \to \infty$ asymptotics (2.13) are fulfilled. Analogously, moving to infinity along the straight line $\mu_2(x + 2\nu_2 y - \tau_2 t) + \delta_2 = 0$, we easily get that at $(\nu_1 - \nu_2)\mu_1 > 0$ our solution has zero asymptotics as $y \to \infty$ and as $y \to -\infty$ asymptotics (2.14) are valid. In the opposite situation, if $(\nu_1 - \nu_2)\mu_1 < 0$, moving to infinity along the straight line mentioned above we see that our solution has zero asymptotics as $y \to -\infty$ and asymptotics (2.14) are valid as $y \to \infty$. Hence it follows that at $(\omega_1 - \omega_2)(\omega_3^3 - \omega_4^3) = 0$ and $\nu_1 \neq \nu_2$ our solution describes the cancellation of one soliton by another. This phenomenon has recently been discovered in another nonlinear integrable system.[3]

Let us analyze now the position of nonzero asymptotics of both the solitons. With (2.15) we get $(\mu_3^2 - 3\nu_3^2)(\mu_4^2 - 3\nu_4^2)\mu_1\mu_2\mu_3\mu_4 > 0$. Then, based on the equality $(\omega_1 - \omega_2)(\omega_3^3 - \omega_4^3) = 0$ we find that $[(\mu_3^2 - 3\nu_3^2)\mu_3 - (\mu_4^2 - 3\nu_4^2)\mu_4](\mu_1 - \mu_2) = 0$. Thus the inequalities $\mu_1\mu_2 > 0$ and $(\mu_3^2 - 3\nu_3^2)(\mu_4^2 - 3\nu_4^2)\mu_3\mu_4 > 0$ are valid. Hence, it follows that in the situation under consideration inequality (2.8) is valid, i.e., the solution considered does not possess singularities at any real values of x, y, and t.

Moreover, by virtue of the inequality $\mu_1\mu_2>0$, the inequality $(\nu_1-\nu_2)\mu_1>0$ results in $(\nu_2-\nu_1)\mu_2<0$. This means that if $(\nu_1-\nu_2)\mu_1>0$, in the upper half-plane, i.e., at

$$y \gg \frac{\mu_2\delta_1-\mu_1}{2(\nu_2-\nu_1)\mu_1\mu_2}+\frac{1}{2}\frac{\tau_2-\tau_1}{\nu_2-\nu_1}t$$

our solution has asymptotics (2.13) and in the lower half-plane, i.e., at

$$y \ll \frac{\mu_1\delta_2-\mu_2}{2(\nu_1-\nu_2)\mu_1\mu_2}+\frac{1}{2}\frac{\tau_1-\tau_2}{\nu_1-\nu_2}t,$$

asymptotics (2.14) hold; vice versa if $(\nu_1-\nu_2)\mu_1<0$, in the upper half-plane, i.e., at

$$y \gg \frac{\mu_1\delta_2-\mu_2}{2(\nu_1-\nu_2)\mu_1\mu_2}+\frac{1}{2}\frac{\tau_1-\tau_2}{\nu_1-\nu_2}t,$$

our solution has asymptotics (2.14) and in the lower half-plane, i.e., at

$$y \ll \frac{\mu_2\delta_1-\mu_1}{2(\nu_2-\nu_1)\mu_1\mu_2}+\frac{1}{2}\frac{\tau_2-\tau_1}{\nu_2-\nu_1}t,$$

asymptotics (2.13) are valid. Thus nonzero asymptotics of both solitons are always on different sides of some straight line parallel to the x axis.

Consider finally the case when $(\omega_1-\omega_2)(\omega_3^3-\omega_4^3)=0$, $\nu_1=\nu_2$, and $\tau_1\neq\tau_2$. Assume that $\nu=\nu_1=\nu_2$. Then at $(\tau_1-\tau_2)\mu_2>0$ as $t\to-\infty$ our solution contains a moving wave of the form

$$\begin{aligned} u&=\frac{2\mu_1^2}{\cosh^2[\mu_1(x+2\nu y-\tau_1 t)+\delta_1]},\\ \varphi&=\hat{a}_1\frac{\exp[i\nu(x+2\nu y)+i\sigma_1 t]}{\cosh[\mu_1(x+2\nu y-\tau_1 t)+\delta_1]}\\ &\quad\times\exp[-i(\mu_1^2+\nu^2)y], \end{aligned} \tag{2.16}$$

where $\delta_1=\frac{1}{2}\ln\alpha_1$, $\hat{a}_1=a_1\exp(-\delta_1)$, and at $(\tau_1-\tau_2)\mu_2<0$ this wave appears at $t\to\infty$. Moreover, at $(\tau_2-\tau_1)\mu_1>0$ as $t\to-\infty$ our solution contains the second moving wave

$$\begin{aligned} u&=\frac{2\mu_2^2}{\cosh^2[\mu_2(x+2\nu y-\tau_2 t)+\delta_2]},\\ \varphi&=\hat{a}_2\frac{\exp[i\nu(x+2\nu y)+i\sigma_2 t]}{\cosh[\mu_2(x+2\nu y-\tau_2 t)+\delta_2]}\\ &\quad\times\exp[-i(\mu_2^2+\nu^2)y], \end{aligned} \tag{2.17}$$

where $\delta_2=\frac{1}{2}\ln\alpha_2$, $\hat{a}_2=a_2\exp(-\delta_2)$ and at $(\tau_2-\tau_1)\mu_1<0$ this wave appears as $t\to\infty$.

By virtue of $\mu_1\mu_2>0$ it follows from the aforesaid that when $(\tau_1-\tau_2)\mu_2>0$ as $t\to-\infty$ our solution has asymptotics (2.16) and as $t\to\infty$ asymptotics (2.17) are valid; and vice versa when $(\tau_1-\tau_2)\mu_2<0$ and $t\to-\infty$ asymptotics (2.17) hold and asymptotics (2.16) hold as $t\to\infty$. Thus in the situation under consideration our solution describes the evolution of soliton (2.16) into soliton (2.17) and vice versa. In the process of evolution the soliton parameters change. By virtue of the equalities

$$4\varkappa|\hat{a}_1|^2=[\tau_1-4(\mu_1^2-3\nu^2)]\mu_1^2,$$
$$4\varkappa|\hat{a}_2|^2=[\tau_2-4(\mu_2^2-3\nu^2)]\mu_2^2,$$

the relations

$$\begin{aligned}\tau_1&=4\mu_1^2+4\varkappa|\hat{a}_1|^2/\mu_1^2-12\nu^2,\\ \tau_2&=4\mu_2^2+4\varkappa|\hat{a}_2|^2/\mu_2^2-12\nu^2\end{aligned} \tag{2.18}$$

are fulfilled, from which it follows that the quantities μ_1^2, μ_2^2, $|\hat{a}_1|$, and $|\hat{a}_2|$ may take any positive values satisfying the condition $(\mu_1-\mu_2)(\mu_2|\hat{a}_1|^2-\mu_1|\hat{a}_2|^2)=0$. Assume that these quantities are chosen so as to fulfill the condition $(\mu_1^4+\varkappa|\hat{a}_1|^2)\mu_2^2\neq(\mu_2^4+\varkappa|\hat{a}_2|^2)\mu_1^2$, i.e., $\tau_1\neq\tau_2$. Then let ζ_1 and ζ_2 be equal to a smaller and a larger of the quantities $\frac{1}{3}(\mu_1^2+\varkappa|\hat{a}_1|^2\mu_1^{-2})$ and $\frac{1}{3}(\mu_2^2+\varkappa|\hat{a}_2|^2\mu_2^{-2})$, respectively. Then from (2.18) it follows that if ν^2 is outside the interval (ζ_1,ζ_2) the inequality $\tau_1\tau_2>0$ holds; otherwise, i.e., at $\nu^2\in(\zeta_1,\zeta_2)$ we have $\tau_1\tau_2<0$. This means that if $\nu^2\bar{\in}(\zeta_1,\zeta_2)$, solitons (2.16) and (2.17) move in the same direction, but if $\nu^2\in(\zeta_1,\zeta_2)$, solitons (2.16) and (2.17) move in opposite directions. Thus at $\nu^2\bar{\in}(\zeta_1,\zeta_2)$ the solution under consideration describes such a transformation of one soliton into another, which does not change the direction of motion. In the opposite situation, at $\nu^2\in(\zeta_1,\zeta_2)$ our solution describes a transformation that changes the direction of the soliton motion to the opposite one as though the soliton is reflected. It is to be noted that the inclusion of $\nu^2\in(\zeta_1,\zeta_2)$ is possible only if $\zeta_2>0$, i.e., in the case when at least one of the quantities $\mu_1^4+\varkappa|\hat{a}_1|^2$ and $\mu_2^4+\varkappa|\hat{a}_2|^2$ is positive.

III. INTERACTION OF TWO WAVES ON THE x AXIS

Now we use the functions $\hat{D}$ and $\hat{\Phi}$ following from the functions D and Φ of the form (2.4) if y is changed by t and t by y. According to (2.6) the functions

$$\hat{u}=2\frac{\partial^2}{\partial x^2}\ln\hat{D},\quad \hat{\varphi}=\frac{\hat{\Phi}}{\hat{D}} \tag{3.1}$$

satisfy the system of equations

$$\begin{aligned} &3\frac{\partial^2\hat{u}}{\partial t^2}-\frac{\partial}{\partial x}\left[\frac{\partial\hat{u}}{\partial y}+\frac{\partial}{\partial x}\left(3\hat{u}^2+\frac{\partial^2\hat{u}}{\partial x^2}+8\varkappa|\hat{\varphi}|^2\right)\right]=0,\\ &i\frac{\partial\hat{\varphi}}{\partial t}=\hat{u}\hat{\varphi}+\frac{\partial^2\hat{\varphi}}{\partial x^2}, \end{aligned} \tag{3.2}$$

resulting from system (2.1) with the same change of variables. Now, in solution (3.1), we change x by $x-4cy$, i.e., assume

$$v(x,y,t)=\hat{u}(x-4cy,y,t),\quad \psi(x,y,t)=\hat{\varphi}(x-4cy,y,t), \tag{3.3}$$

where c is an arbitrary real constant. One can easily be convinced that by virtue of (3.2) functions v and ψ satisfy the system of equations

$$\begin{aligned} &3\frac{\partial^2 v}{\partial t^2}-\frac{\partial^2 v}{\partial x\,\partial y}-\frac{\partial^2}{\partial x^2}\left(4cv+3v^2+\frac{\partial^2 v}{\partial x^2}+8\varkappa|\psi|^2\right)=0,\\ &i\frac{\partial\psi}{\partial t}=v\psi+\frac{\partial^2\psi}{\partial x^2}. \end{aligned} \tag{3.4}$$

It follows from (2.4) that if the conditions

$$\tau_1=4c=\tau_2+4c=0,\quad \sigma_1-4c\nu_1=\sigma_2-4c\nu_2=0 \tag{3.5}$$

are fulfilled, the solution of system (3.4) thus obtained is independent of y, i.e., satisfies the system of equations

$$3\frac{\partial^2 v}{\partial t^2}-\frac{\partial^2}{\partial x^2}\left(4cv+3v^2+\frac{\partial^2 v}{\partial x^2}+8\varkappa|\psi|^2\right)=0,$$

$$\mathrm{i}\frac{\partial\psi}{\partial t}=v\psi+\frac{\partial^2\psi}{\partial x^2}, \tag{3.6}$$

that is important in many branches of mathematical physics.

With (2.5) equalities (3.5) are equivalent to relations $c\omega_1+\omega_1^3+\overline{\omega}_3^3=0$ and $c\omega_2+\omega_2^3+\overline{\omega}_4^3=0$. Using these relations we eliminate the quantities ω_3 and ω_4 from the expressions for the functions $\widehat{D}$ and $\widehat{\Phi}$. As a result, we get functions Δ and Ψ of the form

$$\begin{aligned}\Delta&=1+\alpha_1\exp[2\mu_1(x+2\nu_1 t)]+\alpha_2\exp[2\mu_2(x+2\nu_2 t)]\\&\quad+\gamma_0\exp[2\mu_1(x+2\nu_1 t)+2\mu_2(x+2\nu_2 t)]\\&\quad+2\delta_0\exp[\mu_1(x+2\nu_1 t)+\mu_2(x+2\nu_2 t)]\cos\theta,\\\Psi&=2a_1\{1+\beta_2\exp[2\mu_2(x+2\nu_2 t)]\}\exp[\mu_1(x+2\nu_1 t)]\\&\quad\times\exp[\mathrm{i}\nu_1(x+2\nu_1 t)-\mathrm{i}(\mu_1^2+\nu_1^2)t]\\&\quad+2a_2\{1+\beta_1\exp[2\mu_1(x+2\nu_1 t)]\}\\&\quad\times\exp[\mu_2(x+2\nu_2 t)]\\&\quad\times\exp[\mathrm{i}\nu_2(x+2\nu_2 t)-\mathrm{i}(\mu_2^2+\nu_2^2)t],\end{aligned} \tag{3.7}$$

where

$$\begin{aligned}&\omega_1=\mu_1+\mathrm{i}\nu_1,\quad\omega_2=\mu_2+\mathrm{i}\nu_2,\\&\alpha_1=-\frac{\varkappa|a_1|^2}{(c+\mu_1^2-3\nu_1^2)\mu_1^2},\quad\alpha_2=-\frac{\varkappa|a_2|^2}{(c+\mu_2^2-3\nu_2^2)\mu_2^2},\\&\beta_1=\alpha_1\frac{(\omega_2-\omega_1)^2}{(\omega_2+\overline{\omega}_1)^2}\frac{\omega_2^2+\omega_2\omega_1+\omega_1^2+c}{\omega_2^2-\omega_2\overline{\omega}_1+\overline{\omega}_1^2+c},\quad\beta_2=\alpha_2\frac{(\omega_1-\omega_2)^2}{(\omega_1+\overline{\omega}_2)^2}\frac{\omega_1^2+\omega_1\omega_2+\omega_2^2+c}{\omega_1^2-\omega_1\overline{\omega}_2+\overline{\omega}_2^2+c},\\&\gamma_0=\alpha_1\alpha_2\left|\frac{\omega_1-\omega_2}{\omega_1+\overline{\omega}_2}\right|^4\left|\frac{\omega_1^2+\omega_1\omega_2+\omega_2^2+c}{\omega_1^2-\omega_1\overline{\omega}_2+\overline{\omega}_2^2+c}\right|^2,\\&\delta_0^2=\frac{|\omega_1+\overline{\omega}_1|^2|\omega_2+\overline{\omega}_2|^2|\omega_1^2-\omega_1\overline{\omega}_1+\overline{\omega}_1^2+c|\,|\omega_2^2-\omega_2\overline{\omega}_2+\overline{\omega}_2^2+c|}{|\omega_1+\overline{\omega}_2|^4|\omega_1^2-\omega_1\overline{\omega}_2+\overline{\omega}_2^2+c|^2}|\alpha_1\alpha_2|,\\&\theta=(\nu_2-\nu_1)x+(\mu_1^2-\nu_1^2-\mu_2^2+\nu_2^2)t+\theta_0.\end{aligned} \tag{3.8}$$

By virtue of (3.1) and (3.3) the functions

$$v=2\frac{\partial^2}{\partial x^2}\ln\Delta,\quad\psi=\frac{\Psi}{\Delta} \tag{3.9}$$

satisfy system (3.6). One can easily verify that if the conditions

$$(c+\mu_1^2-3\nu_1^2)\varkappa<0,\quad(c+\mu_2^2-3\nu_2^2)\varkappa<0, \tag{3.10}$$

$$\frac{|\omega_1+\overline{\omega}_1|^2|\omega_2+\overline{\omega}_2|^2|\omega_1^2-\omega_1\overline{\omega}_1+\overline{\omega}_1^2+c|\,|\omega_2^2-\omega_2\overline{\omega}_2+\overline{\omega}_2^2+c|}{|\omega_1+\overline{\omega}_2|^4|\omega_1^2-\omega_1\overline{\omega}_2+\overline{\omega}_2^2+c|^2}\leqslant 1, \tag{3.11}$$

are fulfilled, then using (3.7) and (3.8) one finds that the function Δ is positive at any real values of x and t, and consequently, the solution of system (3.6) defined by (3.9) has no singularities at any real x and t.

Now let us analyze the behavior of this solution. We shall begin with the case when $\omega_1\neq\omega_2$ and $\omega_1^2+\omega_1\omega_2+\omega_2^2+c\neq 0$. Assume that $\nu_1\neq\nu_2$. Then, at $(\nu_2-\nu_1)\mu_2>0$ and $t\to-\infty$ our solution contains a moving wave of the form

$$\begin{aligned}v\sim v_1^-&=\frac{2\mu_1^2}{\cosh^2[\mu_1(x+2\nu_1 t)+\delta_1^-]},\\\psi\sim\psi_1^-&=a_1^-\frac{\exp[\mathrm{i}\nu_1(x+2\nu_1 t)]}{\cosh[\mu_1(x+2\nu_1 t)+\delta_1^-]}\\&\quad\times\exp[-\mathrm{i}(\mu_1^2+\nu_1^2)t],\end{aligned} \tag{3.12}$$

where

$$\delta_1^-=\tfrac{1}{2}\ln\alpha_1,\quad a_1^-=a_1\exp(-\delta_1^-).$$

At $t\to\infty$ this wave has the form

$$\begin{aligned}v\sim v_1^+&=\frac{2\mu_1^2}{\cosh^2[\mu_1(x+2\nu_1 t)+\delta_1^+]},\\\psi\sim\psi_1^+&=a_1^+\frac{\exp[\mathrm{i}\nu_1(x+2\nu_1 t)]}{\cosh[\mu_1(x+2\nu_1 t)+\delta_1^+]}\\&\quad\times\exp[-\mathrm{i}(\mu_1^2+\nu_1^2)t],\end{aligned} \tag{3.13}$$

where

$$\delta_1^+=\tfrac{1}{2}(\ln\gamma_0-\ln\alpha_2),\quad a_1^+=a_1\alpha_2^{-1}\beta_2\exp(-\delta_1^+).$$

By virtue of (3.8) the following equalities are valid:

$$\begin{aligned}\delta_+&=\delta_1^+-\delta_1^-\\&=2\ln\left|\frac{\omega_1-\omega_2}{\omega_1+\overline{\omega}_2}\right|+\ln\left|\frac{\omega_1^2+\omega_1\omega_2+\omega_2^2+c}{\omega_1^2-\omega_1\overline{\omega}_2+\overline{\omega}_2^2+c}\right|,\\|a_1^+|&=|a_1^-|.\end{aligned}$$

Note that if $(\nu_2-\nu_1)\mu_2<0$, then as $t\to-\infty$ the asymptotics of the wave considered have the form (3.13) and as $t\to\infty$ they have the form (3.12). Furthermore, our solution contains the second moving wave. At $(\nu_1-\nu_2)\mu_1>0$ as $t\to-\infty$ it has the form

$$v\sim v_2^-=\frac{2\mu_2^2}{\cosh^2[\mu_2(x+2\nu_2 t)+\delta_2^-]},$$

$$\psi \sim \psi_2^- = a_2^- \frac{\exp[i\nu_2(x+2\nu_2 t)]}{\cosh[\mu_2(x+2\nu_2 t)+\delta_2^-]} \times \exp[-i(\mu_2^2+\nu_2^2)t], \quad (3.14)$$

where

$$\delta_2^- = \tfrac{1}{2}\ln\alpha_2, \quad a_2^- = a_2\exp(-\delta_2^-),$$

and as $t\to\infty$ the asymptotics of this wave have the form

$$v \sim v_2^+ = \frac{2\mu_2^2}{\cosh^2[\mu_2(x+2\nu_2 t)+\delta_2^+]},$$

$$\psi \sim \psi_2^+ = a_2^+ \frac{\exp[i\nu_2(x+2\nu_2 t)]}{\cosh[\mu_2(x+2\nu_2 t)+\delta_2^+]} \times \exp[-i(\mu_2^2+\nu_2^2)t], \quad (3.15)$$

where

$$\delta_2^+ = \tfrac{1}{2}(\ln\gamma_0 - \ln\alpha_1), \quad a_2^+ = a_2\alpha_1^{-1}\beta_1\exp(-\delta_2^+).$$

With (3.8) we get that

$$\delta_2 = \delta_2^+ - \delta_2^- = 2\ln\left|\frac{\omega_1-\omega_2}{\omega_1+\overline{\omega}_2}\right| + \ln\left|\frac{\omega_1^2+\omega_1\omega_2+\omega_2^2+c}{\omega_1^2-\omega_1\overline{\omega}_2+\overline{\omega}_2^2+c}\right|,$$

$$|a_2^+| = |a_2^-|.$$

It is obvious that if $(\nu_1-\nu_2)\mu_1<0$, asymptotics of the second wave as $t\to-\infty$ have the form (3.15) and as $t\to\infty$ they have the form (3.14).

Thus, at $\omega_1\neq\omega_2$, $\omega_1^2+\omega_1\omega_2+\omega_2^2+c\neq 0$, and $\nu_1\neq\nu_2$ the solution under consideration describes the interaction of two waves. The nonlinear nature of the interaction leads to the distortion of both the waves in the vicinity of the point $x=x_0$, where

$$x_0 = \frac{(\delta_2^+ + \delta_2^-)\mu_1\nu_1 - (\delta_1^+ + \delta_1^-)\mu_2\nu_2}{2(\nu_2-\nu_1)\mu_1\mu_2}.$$

This distortion becomes maximal at the time moment $t=t_0$, where

$$t_0 = \frac{(\delta_1^+ + \delta_1^-)\mu_2 - (\delta_2^+ + \delta_2^-)\mu_1}{4(\nu_2-\nu_1)\mu_1\mu_2},$$

and tends to zero as $t\to\pm\infty$. Far from the point $x=x_0$, $t=t_0$ each of the waves has a form given by one of the asymptotics (3.12)–(3.15). The interaction gives exceptional results in the phase shifts of both the waves, i.e., the interaction is elastic.

In the case when $\omega_1\neq\omega_2$, $\omega_1^2+\omega_1\omega_2+\omega_2^2+c\neq 0$, and $\nu_1=\nu_2$ the solution under consideration describes one solitary wave formed by two merged waves that move as one. However, the shape of this wave differs strongly from that of the initial waves. For instance, let $\omega_1=2\mu+i\nu$ and $\omega_2=\mu+i\nu$, where the quantities μ and ν satisfy the condition $\mu^2-3\nu^2+c=0$. Assuming then that $a_2=0$, one can easily find that the functions Δ and Ψ in this case are

$$\Delta = 1+\alpha_2\exp[2\mu(x+2\nu t)]+\alpha_1\exp[4\mu(x+2\nu t)] + \tfrac{1}{9}\alpha_1\alpha_2\exp[6\mu(x+2\nu t)],$$

$$\Psi = 2a_1\{1+\tfrac{1}{3}\alpha_2\exp[2\mu(x+2\nu t)]\}\exp[2\mu(x+2\nu t)] \times\exp[i\nu(x+2\nu t)-i(4\mu^2+\nu^2)t],$$

where $\alpha_1 = -\varkappa|a_1|^2/12\mu^4$ and α_2 can be chosen arbitrarily. Assuming $\alpha_2=(3\alpha_1)^{1/2}$ we find that

$$\Delta = \{1+(\alpha_1/3)^{1/2}\exp[2\mu(x+2\nu t)]\}^3.$$

At $\varkappa=-1$ and $a_1\neq 0$ we have $\alpha_1>0$, and consequently, $\Delta>1$ at any real values of x and t. Thus, the solution under consideration has the form

$$v = \frac{6\mu^2}{\cosh^2[\mu(x+2\nu t)+\delta]},$$

$$\psi = a\frac{\exp[i\nu(x+2\nu t)]}{\cosh^2[\mu(x+2\nu t)+\delta]}\exp[-i(4\mu^2+\nu^2)t],$$

where

$$\delta = \tfrac{1}{4}\ln(\alpha_1/3) = \tfrac{1}{4}\ln(|a_1|/6\mu^2), \quad a = \tfrac{1}{2}a_1\exp(-2\delta)$$

and consequently, the equality $|a|=3\mu^2$ is valid.

Finally, at $\omega_1=\omega_2$ our solution degenerates into a one-soliton solution.

Now consider the case when $\omega_1\neq\omega_2$ but $\omega_1^2+\omega_1\omega_2+\omega_2^2+c=0$. Assume that

$$\omega_2^+ = -\omega_1/2+(i/2)(3\omega_1^2+4c)^{1/2}, \quad (3.16)$$

$$\omega_2^- = -\omega_1/2-(i/2)(3\omega_1^2+4c)^{1/2}. \quad (3.17)$$

Now we use the complex plane ω_1 and remove from it a piece of the straight line connecting the points $\omega_1^+=2i(c/3)^{1/2}$ and $\omega_1^-=-2i(c/3)^{1/2}$ of the imaginary axis if $c>0$ and the points $\omega_1^+=2(-c/3)^{1/2}$ and $\omega_1^-=-2(-c/3)^{1/2}$ of the real axis if $c<0$. In the complex plane ω_1 with such a cut the equalities (3.16) and (3.17) define the mappings F_+ and F_- into the complex plane ω_2. It follows from (3.16) that at $c\neq 0$ and at large $|\omega_1|$ the mapping F_+ is close to the rotation by the angle $2\pi/3$ and from (3.17) it follows that at $c\neq 0$ and at large $|\omega_1|$ the mapping F_- is close to the rotation by the angle $-2\pi/3$. At $c=0$ these mappings are simply rotations by the angles $2\pi/3$ and $-2\pi/3$, respectively. Now in the complex plane ω_1 with the above cut we take the regions G_+ and G_- defined by the inequalities $(\mu_1^2-3\nu_1^2+c)\mu_1>0$ and $(\mu_1^2-3\nu_1^2+c)\mu_1<0$, respectively. One can easily see that the region G_+ consists of three components H_1, H_3, and H_5, and the region G_- consists of three components H_2, H_4, and H_6 determined as follows:

$$H_1=\{(\mu_1,\nu_1):\ \mu_1>0,\ \mu_1^2-3\nu_1^2+c>0\},$$
$$H_2=\{(\mu_1,\nu_1):\ \mu_1>0,\ \nu_1>0,\ \mu_1^2-3\nu_1^2+c<0\},$$
$$H_3=\{(\mu_1,\nu_1):\ \mu_1<0,\ \nu_1>0,\ \mu_1^2-3\nu_1^2+c<0\},$$
$$H_4=\{(\mu_1,\nu_1):\ \mu_1<0,\ \mu_1^2-3\nu_1^2+c>0\},$$
$$H_5=\{(\mu_1,\nu_1):\ \mu_1<0,\ \nu_1<0,\ \mu_1^2-3\nu_1^2+c<0\},$$
$$H_6=\{(\mu_1,\nu_1):\ \mu_1>0,\ \nu_1<0,\ \mu_1^2-3\nu_1^2+c<0\}.$$

Let $\Gamma_1,\ldots,\Gamma_6$ be boundaries of the regions $H_1,\ldots,H_6$, respectively. By using (3.16) and (3.17) one can easily be convinced that the equalities

$$F_+(\Gamma_1)=\Gamma_3, \quad F_+(\Gamma_2)=\Gamma_4,$$
$$F_+(\Gamma_3)=\Gamma_5, \quad F_+(\Gamma_4)=\Gamma_6,$$
$$F_+(\Gamma_5)=\Gamma_1, \quad F_+(\Gamma_6)=\Gamma_2,$$
$$F_-(\Gamma_1)=\Gamma_5, \quad F_-(\Gamma_2)=\Gamma_6,$$

$$F_-(\Gamma_3)=\Gamma_1,\quad F_-(\Gamma_4)=\Gamma_2,$$
$$F_-(\Gamma_5)=\Gamma_3,\quad F_-(\Gamma_6)=\Gamma_4$$

are valid. On the basis of the well-known theorems of the theory of conformal mappings there follow analogous relations between the regions $H_1,...,H_6$ and their images for the mappings F_+ and F_-, i.e.,

$$F_+(H_1)=H_3,\quad F_+(H_2)=H_4,$$
$$F_+(H_3)=H_5,\quad F_+(H_4)=H_6,$$
$$F_+(H_5)=H_1,\quad F_+(H_6)=H_2,$$
$$F_-(H_1)=H_5,\quad F_-(H_2)=H_6,$$
$$F_-(H_3)=H_1,\quad F_-(H_4)=H_2,$$
$$F_-(H_5)=H_3,\quad F_-(H_6)=H_4.$$

In the complex plane ω_1 with the above cut we now take the regions H_+ and H_- defined by the inequalities $(\mu_1^2-3\nu_1^2+c)\mu_1^2>0$ and $(\mu_1^2-3\nu_1^2+c)\mu_1^2<0$, respectively. It is easily seen that $H_+=H_1\cup H_4$ and $H_-=H_2\cup H_3\cup H_5\cup H_6$. It follows from that aforesaid that $F_+(H_+)\cap H_+=F_-(H_+)\cap H_+=0$, and vice versa $F_+(H_-)\cap H_-=H_2\cup H_5$, $F_-(H_-)\cap H_-=H_3\cup H_6$. This implies that whatever point $\omega_1\in H_+$ might be used, in the region H_+ there is no point ω_2 that would satisfy the relation $\omega_1^2+\omega_1\omega_2+\omega_2^2+c=0$. On the contrary, for any point $\omega_1\in H_-$ there will be only one point $\omega_2\in H_-$ and that $\omega_1^2+\omega_1\omega_2+\omega_2^2+c=0$. In this case the following relations occur:

(1) if $\omega_1\in H_2$, then $\omega_2=F_-(\omega_1)\in H_6$,

(2) if $\omega_1\in H_3$, then $\omega_2=F_+(\omega_1)\in H_5$,

(3) if $\omega_1\in H_5$, then $\omega_2=F_-(\omega_1)\in H_3$,

(4) if $\omega_1\in H_6$, then $\omega_2=F_+(\omega_1)\in H_2$.

This implies that the real μ_1,μ_2 and imaginary ν_1,ν_2 parts of the points ω_1 and ω_2 satisfy the conditions $\mu_1\mu_2>0$ and $\nu_1\nu_2<0$. According to (3.8) it follows that at $\varkappa=1$ the inequalities $\alpha_1\geqslant 0$, $\alpha_2\geqslant 0$, and $0\leqslant\delta_0^2\leqslant\alpha_1\alpha_2$ are valid. Thus in the considered case our solution has no singularities at any real values of x and t.

Let us analyze the behavior of this solution. It is easily seen that at $(\nu_2-\nu_1)\mu_2>0$ as $t\to-\infty$ our solution has a moving wave of the form

$$v=\frac{2\mu_1^2}{\cosh^2[\mu_1(x+2\nu_1t)+\delta_1]},$$
$$\psi=\hat{a}_1\frac{\exp[i\nu_1(x+2\nu_1t)]}{\cosh[\mu_1(x+2\nu_1t)+\delta_1]}\times\exp[-i(\mu_1^2+\nu_1^2)t],\qquad(3.18)$$

where $\delta_1=\frac{1}{2}\ln\alpha_1$, $\hat{a}_1=a_1\exp(-\delta_1)$, and at $(\nu_2-\nu_1)\mu_2<0$ this wave appears as $t\to\infty$. Moreover, at $(\nu_1-\nu_2)\mu_1>0$ as $t\to-\infty$ our solution has the second moving wave

$$v=\frac{2\mu_2^2}{\cosh^2[\mu_2(x+2\nu_2t)+\delta_2]},$$
$$\psi=\hat{a}_2\frac{\exp[i\nu_2(x+2\nu_2t)]}{\cosh[\mu_2(x+2\nu_2t)+\delta_2]}\times\exp[-i(\mu_2^2+\nu_2^2)t],\qquad(3.19)$$

where $\delta_2=\frac{1}{2}\ln\alpha_2$, $\hat{a}_2=a_2\exp(-\delta_2)$, and at $(\nu_1-\nu_2)\mu_1<0$ this wave appears as $t\to\infty$. By virtue of $\mu_1\mu_2>0$ from $(\nu_2-\nu_1)\mu_2>0$ there follows the inequality $(\nu_1-\nu_2)\mu_1<0$. This means that if $(\nu_2-\nu_1)\mu_2>0$, then as $t\to-\infty$ our solution has asymptotics (3.18) and as $t\to\infty$ it has asymptotics (3.19). On the contrary, if $(\nu_2-\nu_1)\mu_2<0$, then as $t\to-\infty$ asymptotics (3.19) are valid whereas asymptotics (3.18) hold as $t\to\infty$. Thus in the considered situation our solution has essentially different asymptotics as $t\to-\infty$ and $t\to\infty$. Then, by virtue of the inequality $\nu_1\nu_2<0$ the moving waves (3.18) and (3.19) have opposite directions of motion, i.e., at $(\nu_2-\nu_1)\mu_2>0$ soliton (3.18) comes from infinity, then its parameters change, it changes the direction of its motion to an opposite one, and finally, goes to infinity in the form (3.19). At $(\nu_2-\nu_1)\mu_2<0$ the process proceeds in the opposite order, i.e., soliton (3.19) comes from infinity and soliton (3.18) goes to infinity. Note that according to (3.8) the following relation is valid:

$$\frac{|\hat{a}_1|^2}{|\hat{a}_2|^2}=\frac{(\mu_1^2-3\nu_1^2+c)\mu_1^2}{(\mu_2^2-3\nu_2^2+c)\mu_2^2}=\frac{\mu_1}{\mu_2}.$$

In conclusion we should like to note that solutions analogous to those considered above have earlier been found in another system of equations.[4]

[1]V. K. Mel'nikov, Lett. Math. Phys. 7, 129 (1983).
[2]V. K. Mel'nikov, JINR preprint P2-86-689, Dubna, 1986.
[3]V. K. Mel'nikov, JINR preprint P2-86-234, Dubna, 1986.
[4]F. Calogero and A. Degasperis, Lett. Nuovo Cimento 16, 425 (1976).

Yang–Mills solutions in $S^3 \times S^1$

Javier Villarroel
Departamento de Física Teórica, Facultad de Ciencias, Universidad de Salamanca, 37008-Salamanca, Spain

(Received 16 September 1986; accepted for publication 27 May 1987)

A search is conducted for new solutions to Yang–Mills classical field theory defined on $S^3 \times S^1$ using Witten's *Ansatz*. Then some sort of non-Abelian plane waves giving rise to a divergent energy are obtained.

I. INTRODUCTION

The interest in looking for new solutions of the classical Yang–Mills field equations arises from the hope that a greater knowledge of the classical theory will help us to understand the quantal aspects. Many interesting solutions have been found (see Actor[1] for a review), but we think also that its study needs to continue. One of the more celebrated *Ansätze* was given by Witten,[2] who was also able to solve the self-dual SU(2) equation derived from it, obtaining then a multi-instanton solution. While Witten exposed it in Euclidean space-time, we shall use the corresponding Minkowski version of it, with the obvious changes when necessary, and we refer to Witten's paper[2] for the notation.

On the other hand, the pure Yang–Mills theory is invariant under the full conformal group SO(4,2), and we can use this fact in order to obtain further information. In particular, $O(4) \times O(2)$, the largest compact subgroup of SO(4,2), has revealed special interest and has led many authors to the discovery of new solutions (see Refs. 3 and 4). We shall profit from this formalism by projecting the SU(2) Yang–Mills theory on the hypertorus, and then by using Witten's *Ansatz*, we shall try to find solutions to the general non-self-dual field equations. As many authors have described in detail how this process works (see Refs. 5–7), we shall not repeat the procedure, but in order to fix the notation it is worth noticing that we define the nontrivial hypertorus angles ω, $\tilde\omega$ as

$$\omega = \arctan[2t/(1+r^2-t^2)],$$
$$\tilde\omega = \arctan[2r/(1+t^2-r^2)].$$

As in Witten's *Ansatz* we restrict ourselves to (t,r) coordinates. We only need to consider the angles ω, $\tilde\omega$ to describe the situation.

II. THE SOLUTIONS

Unlike Witten, we are interested in non-self-dual second-order equations that follow from his *Ansatz*

$$\partial_t(r^2 f_{01}) = 2(\phi_1 D_r \phi_2 - \phi_2 D_r \phi_1),$$
$$\partial_r(r^2 f_{01}) = 2(\phi_1 D_t \phi_2 - \phi_2 D_t \phi_1),$$
$$r^2(D_t D_t - D_r D_r)\phi_a = \phi_a(1-\phi_1^2-\phi_2^2), \quad a=1,2,$$

into the new coordinates $(\omega,\tilde\omega)$. We fix the gauge by the condition $\partial^\mu a_\mu = 0$, which is easily seen to be fulfilled by the choice $a_\mu = \chi(\tilde\omega)\partial_\mu \omega$, where $\chi(\tilde\omega)$ is a function to be determined, $a_\mu \equiv (a_0, a_1)$, and we do not restrict $\phi_1(\omega,\tilde\omega)$, $\phi_2(\omega,\tilde\omega)$ for the moment. Then, after some manipulations we obtain for the motion equations

$$\sin^2\tilde\omega\, \chi_{\tilde\omega\tilde\omega} + 2\sin\tilde\omega\cos\tilde\omega\, \chi_{\tilde\omega} + 2\{\phi_1(\phi_2)_\omega - \phi_2(\phi_1)_\omega - \chi(\phi_1^2+\phi_2^2)\} = 0,$$
$$\phi_1(\phi_2)_{\tilde\omega} - \phi_2(\phi_1)_{\tilde\omega} = 0,$$
$$(\phi_1)_{\omega\omega} - (\phi_1)_{\tilde\omega\tilde\omega} + 2\chi(\phi_2)_\omega - \chi^2\phi_1 = \phi_1(1-\phi_1^2-\phi_2^2)/\sin^2\tilde\omega,$$
$$(\phi_2)_{\omega\omega} - (\phi_2)_{\tilde\omega\tilde\omega} - 2\chi(\phi_1)_\omega - \chi^2\phi_2 = \phi_2(1-\phi_1^2-\phi_2^2)/\sin^2\tilde\omega.$$

The second equation can actually be integrated giving the relation $\phi_2(\omega,\tilde\omega) = \alpha(\omega)\phi_1(\omega,\tilde\omega)$, where α is an arbitrary function of his argument. Although it is possible to simplify the equations, they are still difficult to solve. We shall concentrate henceforth on those fields fulfilling the requirement that α be a constant. Then demanding consistency to the full set we obtain that the following relation must be fulfilled:

$$\chi(\phi_1)_\omega(\alpha^2+1) = 0.$$

(We shall henceforth drop the index and write $\phi_1 \equiv \phi$.) Consider first the case when $\chi = 0$, and also $\alpha = 0$. By substituting we obtain an equation that admits the solution

$$\chi = \phi_2 = 0, \quad \phi = \cos\tilde\omega.$$

But this is not anything new. It is just the meron–antimeron solution of De Alvaro, Fubini, and Furlan but in a different gauge (see Refs. 8 and 9). We must then look for more original solutions.

Now we shall concentrate on the former equations when $\chi,\phi \neq 0$ and $\alpha^2+1=0$. With these conditions we obtain

$$\sin^2\tilde\omega\, \chi_{\tilde\omega\tilde\omega} + 2\sin\tilde\omega\cos\tilde\omega\, \chi_{\tilde\omega} = 0,$$
$$\phi_{\omega\omega} - \phi_{\tilde\omega\tilde\omega} + 2i\chi\phi_\omega - \chi^2\phi - \phi/\sin^2\tilde\omega = 0.$$

We can immediately integrate the first equation giving

$$\chi = Q\cot\tilde\omega + Q_2,$$

where Q, Q_2 are integration constants. The second one is then

$$\phi_{\omega\omega} - \phi_{\tilde\omega\tilde\omega} + \{2iQ\cot\tilde\omega + 2iQ_2\}\phi_\omega - \{Q^2\cot^2\tilde\omega + 2QQ_2\cot\tilde\omega + Q_2^2 + 1/\sin^2\tilde\omega\}\phi = 0.$$

In order to solve this partial differential equation we shall assume that $\phi(\omega,\tilde\omega) = A(\omega)B(\tilde\omega)$, where A, B are unknown functions. When substituting such an expression carefully, we realize that we have to consider the two cases $Q=0$, $Q \neq 0$ separately because the solution corresponding to the first case is more general than that corresponding to the second one when taking the $Q\to 0$ limit, and can never be reached from the general case.

 0022-2488/87/112610-04$02.50

(a) $Q = 0$, then we have that A,B must verify

$$A_{\omega\omega} + 2iQ_2 A_\omega - \mu A = 0,$$

$$B_{\tilde\omega\tilde\omega} + \{Q_2^2 + 1/\sin^2\tilde\omega - \mu\}B = 0,$$

when μ is an arbitrary constant. By solving the first equation we find that in order for the solution to be a global one (that is to say, periodic), we must have

$$Q_2 = n, \quad \mu = n^2 - m^2,$$

where n,m are arbitrary integer numbers. The general solution to the field equations is then

$$\phi = \sum_{m=-\infty}^{\infty} \{C_{1,m} e^{i(m-n)\omega} + C_{2,m} e^{-i(m+n)\omega}\}$$
$$\times [-\sin\tilde\omega]^\alpha \{C_{3,m}\mathbb{F}(a,b,\tfrac{1}{2}, -\cos\tau\omega)$$
$$+ C_{4,m}\{-\cos\tilde\omega\}^{1/2}\mathbb{F}(a+\tfrac{1}{2}, b+\tfrac{1}{2}, \tfrac{3}{2}, -\cos\omega)\},$$

where $\mathbb{F}(a,b,c)$ is the hypergeometric function, and

$$\alpha = (1+\sqrt{3}i)/4, \quad a = \alpha + m/2, \quad b = \alpha - m/2.$$

We see that for fixed m, ϕ depends on two independent functions of ω. Curiously enough such a thing does not happen in the general case.

(b) $Q \neq 0$, looking again for a solution in the way $\phi(\omega,\tilde\omega) = A(\omega)B(\tilde\omega)$ we find that A must verify

$$A\omega = \mu A,$$

where μ is an arbitrary constant. But in order for the solution to be periodic we impose $\mu = im$, m being any integer. Moreover, we have the following equation for B:

$$B\tilde\omega\tilde\omega + \{Q^2\cot^2\tilde\omega + 2Q(Q_2+m)\cot\tilde\omega$$
$$+ (Q_2+m)^2 + (1/\sin^2\tilde\omega)\}B = 0.$$

Before considering the solution to this equation, we shall investigate first the case when $Q_2 + m = 0$. The equation now becomes simplified and we can write its solution as follows:

$$\phi(\omega,\tilde\omega) = e^{-iQ_2\omega}\sin^{2\nu}\tilde\omega\{C_1\mathbb{F}(\alpha,\beta,\gamma,\sin^2\tilde\omega)$$
$$+ C_2(\sin^2\tilde\omega)^{1-\gamma}$$
$$\times\mathbb{F}(\alpha+1-\gamma,\beta+1-\gamma,2-\gamma,\sin^2\tilde\omega)\},$$

where

$$\nu = [1+\sqrt{-(3+4Q^2)}]/2, \quad Q_2\in\mathbb{Z},$$
$$\alpha = (2\nu + iQ)/2,$$
$$\beta = (2\nu - iQ)/2,$$
$$\gamma = 2\nu + \tfrac{1}{2}.$$

Since $\operatorname{Re}(\alpha+\beta-\gamma) = -\tfrac{1}{2}$, we can say that our solution is convergent on the whole disk $|y| \leqslant 1$, with $y = \sin^2\tilde\omega$, that is to say, it converges for every value of $\tilde\omega$: $0 \leqslant \tilde\omega \leqslant \pi$. This is a pleasant property of this solution that together with its relative simplicity has led us to consider it separately. We now shall proceed to investigate the general case, namely when $\eta \equiv Q_2 + m \neq 0$. In order to solve the subsequent equation we carry out the following change of variables:

$$\cot\tilde\omega = i(1-2y),$$
$$B(\tilde\omega) = \exp\{-a[\arctan(i(1-2y))]\}\{4y(y-1)\}^b f(y),$$

and then we can reduce the equation to a hypergeometric form provided we choose the parameters a,b as

$$a = i\eta, \quad b = iQ.$$

The solution then turns out to be

$$\phi = \sum_{m=-\infty}^{\infty} e^{im\omega}e^{-i\eta(\tilde\omega-\pi/2)}\Big\{C_{1,m}\mathbb{F}\Big(\alpha,\beta,\gamma,\frac{1+i\cot\tilde\omega}{2}\Big)$$
$$+ C_{2,m}\Big(\frac{e^{i\tilde\omega}}{2\sin\tilde\omega}\Big)^{-iQ-\eta}$$
$$\times\mathbb{F}\Big(1+\alpha-\gamma,1+\beta-\gamma,2-\gamma,\frac{1+i\cot\tilde\omega}{2}\Big)\Big\},$$

where $\mathbb{F}$ is the hypergeometric function of parameters

$$\alpha = (1+2iQ+\sqrt{-3-4Q^2})/2,$$
$$\beta = (1+2iQ-\sqrt{-3-4Q^2})/2,$$
$$\gamma = 1+\eta+iQ.$$

Some comments are now in order. First, we observe that the superposition, in the index m, of solutions is also a solution. In the second place, we note that in order for our solution to be convergent we must request that $\operatorname{Re}(iQ-\eta) < 0$. Finally, it is worth noticing that under the simultaneous changes $\eta \to -iQ$ and $Q \to i\eta$ we obtain another solution, although with similar properties to the former.

In the conditions already stated, our solutions are convergent in the region $\pi/4 \leqslant \tilde\omega \leqslant 3\pi/4$. As the angle $\tilde\omega$ sweeps over the region $0 \leqslant \tilde\omega \leqslant \pi$, we must also give the form of the solution on the excluded regions. We do it by means of analytic prolongation and we obtain

$$\phi = \sum_{m=-\infty}^{\infty}\Big\{C_{1,m}\frac{\Gamma(\gamma)\Gamma(\beta-\alpha)}{\Gamma(\beta)\Gamma(\gamma-\alpha)} + C_{2,m}\frac{\Gamma(2-\gamma)\Gamma(\beta-\alpha)}{\Gamma(1+\beta-\gamma)\Gamma(1-\alpha)}\Big\}$$
$$\times e^{im\omega}e^{-i\eta\tilde\omega}\Big(\frac{e^{i\tilde\omega}}{2\sin\tilde\omega}\Big)^{-\alpha}\mathbb{F}\Big(\alpha,\alpha+1-\gamma,\alpha+1-\beta,\frac{2\sin\tilde\omega}{e^{i\tilde\omega}}\Big)$$
$$+\sum_{m=-\infty}^{\infty}\Big\{C_{1,m}\frac{\Gamma(\gamma)\Gamma(\alpha-\beta)}{\Gamma(\alpha)\Gamma(\gamma-\beta)} + C_{2,m}\frac{\Gamma(\alpha-\beta)\Gamma(2-\gamma)}{\Gamma(1+\alpha-\gamma)\Gamma(1-\beta)}\Big\}e^{im\omega}e^{-i\eta\tilde\omega}$$
$$\times\Big(\frac{e^{i\tilde\omega}}{2\sin\tilde\omega}\Big)^{-\beta}\mathbb{F}\Big(\beta,1+\beta-\gamma,\beta+1-\alpha,\frac{2\sin\tilde\omega}{e^{i\tilde\omega}}\Big).$$

This prolongation is valid, strictly speaking, in the regions $0<\tilde{\omega}\leqslant\pi/4$, $3\pi/4\leqslant\tilde{\omega}<\pi$. As we also want to include the point $\tilde{\omega}=0$, which corresponds in Minkowski space to the origin of coordinates $r=0$, we must study what happens at that point. We find that for Q real we can extend the solution to this point without problems while for Q imaginary pure $Q=i\nu$, $\nu\in\mathbb{R}$ the solution is finite only if $-1\leqslant\nu\leqslant1$. Finally, if $Q=\rho+i\nu$ we obtain the same condition. We shall see that the physically interesting cases are just those corresponding to Q real or Q imaginary pure with $|Q|\leqslant1$, so in this respect the behavior of our solution is the right one.

It is interesting now to see when our solutions are self-dual. It can be seen that these kinds of solutions must verify

$$D_r\phi=-D_t\phi$$

and also necessarily $Q=i$. Solving the former equation we obtain as the general solution

$$\phi=\sin\tilde{\omega}\, e^{-Q_2\tilde{\omega}i}H(\omega-\tilde{\omega}),$$

where $H(\omega-\tilde{\omega})$ is an arbitrary function. But we must impose also a periodicity condition in the ω angle, that is to say $H(\omega-\tilde{\omega})=H(\omega-\tilde{\omega}+2\pi)$. Then we can develop it in the Fourier series as

$$\phi=\sum_{m=-\infty}^{\infty}a_m e^{im(\omega-\tilde{\omega})}\sin\tilde{\omega}\,e^{-Q_2\tilde{\omega}i}$$

$$=\sum_{m=-\infty}^{\infty}a_m e^{im\omega}e^{-i\eta\tilde{\omega}}\sin\tilde{\omega},$$

which is the general form of a self-dual solution in our particular *Ansatz*.

On the other hand, in the general case, when $Q=i$ the constants α,β are integers so that the hypergeometric series is now degenerate. Using Gauss recurrence relations we obtain that the solution to the second-order equations is then

$$\phi=\sum_{m=-\infty}^{\infty}e^{im\omega}e^{-i\eta\tilde{\omega}}\left\{a_m+b_m\int\frac{e^{2i\eta\tilde{\omega}}}{\sin^2\tilde{\omega}}\,d\tilde{\omega}\right\}\sin\tilde{\omega}.$$

We can identify the first term with the self-dual solution, while the second one does not have anything to do with these kinds of solutions. So, summing up, we can say that for our particular *Ansatz* we obtain self-dual solutions only when $Q=i$, and in this case only some special cases are self-dual (when $b_m=0$).

We now turn to the description of some physical properties of our solutions. First we recall that our solutions are complex.

It is known[10,11] that a theory with gauge group G and complex potential is equivalent to a theory with real potential but in which the gauge group is now the complex extension of G. In our case, it would be SU$(2,C)$. In these references, it is also shown that in the SU$(2,C)$ theory the energy is not positive definite. We shall check it for our particular solutions, and we shall see shortly that under some conditions we can get a real and positive energy momentum tensor. This fact, together with the observation that the only gauge invariant property of our solution $\phi(\omega,\tilde{\omega})$ is its modulus, lets us think that our solutions are physically acceptable (although we do not claim that they have a clear physical meaning). On the other hand, it is known that a self-dual solution in Minkowski space-time[12] or also the Minkowskian solutions obtained through the Φ^4 *Ansatz* are necessarily complex,[13] so it is not too strange that our solutions are also complex. We now go on to consider the form of the electromagnetic potentials A_i^a in terms of the solutions we have just obtained,

$$A_0^a=-\frac{x_a}{r}\left\{\frac{Q(1+t^2)^2-Qr^4}{r[(1+t^2+r^2)^2-4r^2t^2]}+\frac{2Q_2(1+t^2+r^2)}{r[(1+t^2+r^2)^2-4r^2t^2]}\right\},$$

$$A_i^a=\left\{\frac{1}{r}\left(\delta_{ai}-\frac{x_ix_a}{r^2}\right)+i\epsilon_{iaj}\frac{x_j}{r^2}\right\}\phi(r,t)+\frac{\epsilon_{iaj}x_j}{r^2}-\frac{x_ix_a}{r^2}\left\{\frac{2Qt(1+t^2-r^2)}{(1+t^2+r^2)^2-4r^2t^2}+\frac{4Q_2rt}{(1+r^2+t^2)^2-4r^2t^2}\right\}.$$

It is also interesting to know the expressions for the electric and magnetic fields, but as they get very cumbersome we shall only give them for the case of self-dual fields in which case they are much more simple. We have

$$B_j^a=iE_j^a,\quad j=1,2,3,\quad a=1,2,3,$$

$$E_i^a=-\frac{ix_ix_a}{r^4}+2\left\{\frac{1}{r}\left(\delta_{ia}-\frac{x_ix_a}{r^2}\right)+\frac{\epsilon_{iaj}x_j}{r^2}\right\}$$

$$\times\left\{\frac{[(1+t^2-r^2)-2iQ_2]2H(\omega-\tilde{\omega})+2rH'(\omega-\tilde{\omega})}{\{1+(t+r)^2\}^{1/2}\{1+(t-r)^2\}^{3/2}}\right\}\exp\left\{-iQ_2\arctan\frac{2r}{1+t^2-r^2}\right\}.$$

For the components of the energy momentum tensor we obtain

$$\theta_{00}=(1+Q^2)/2r^4,$$

$$\theta_{ij}=(\delta_{ij}-x_ix_j/r^2)(1+Q^2)/2r^4.$$

Let us now study when the energy momentum tensor is positive definite. Clearly it is true for any real value of Q, 0 even for Q imaginary pure, $Q=i\nu$ with $|\nu|\leqslant1$. We recall that in these cases the solution was not divergent in any place so clearly that we have to restrict ourselves to this range of values of the parameter. That is to say, the only solutions which can be physically acceptable are those for which Q is real or even better, where $Q = iv$ with $|v| \leqslant 1$ (in which case the energy gets diminished, and so the solution is more stable), because under these conditions the energy density is positive, and our solution can be defined over the whole hypertorus.

Other quantities of interest are just the pseudoscalar density and the action

$$\mathscr{D} \equiv E_i^a B_i^a = -Q/r^4, \quad \mathscr{L} = (1 - Q^2)/2r^4.$$

We see that in most of the cases these quantities are divergent, although in some situations they are not. In particular, when $Q = i$ the energy is zero, which is not surprising as long as the self-dual solutions correspond just to this case.

Generally speaking, we can say that our solutions represent some sort of plane waves with strong traces of nonlinearity. (This is because we obtain them from an equation in which the fields ϕ and χ were coupled. Note also that a superposition in Q, Q_2 of solutions is not a solution.) It does not look like they have any relation with Coleman's plane waves. Anyway, except in the case $Q = i$, they carry a divergent energy, and in general also an infinite action. They are not confined to any region of the space, so they do not have any similarity with other solutions as instantons, merons, monopoles, or solitons. In fact, we think that although physically acceptable, our solutions have not the pleasant features of the solutions just named, so their physical interpretation is not very clear. But if we consider instead the equivalent problem of the Abelian Higgs model in a two-dimensional space-time[2] we can give a physical meaning to these solutions: They describe how an electric pole of charge Q, located at the origin is interacting with a charged field, namely the ϕ_i, giving then rise to some complicated waves, represented here in terms of hypergeometric functions. The interaction is due to the curvature of the manifold in which they are living, corresponding to a nontrivial metric tensor $g_{00} = -g_{11} = r^2$, $g_{01} = 0$.

III. CONCLUSIONS

We have given a set of, in general, non-self-dual solutions to Yang–Mills field equations using the invariance of this theory under the conformal group. These solutions are not localized and in some sense are some sort of non-Abelian color plane waves, then giving rise to a divergent energy. Then, they do not share the properties that can lead us to interpret them as a localized particle, a fact from which their physical meaning is not transparent. Anyway, we think that it is always worth knowing a little bit more of the Yang–Mills classical theory in order to understand the subsequent quantized theory, particularly in the nonperturbative sector.

ACKNOWLEDGMENT

It is a pleasure for me to thank Professor José M. Cerveró for suggesting this problem to me and also for several interesting discussions during its realization.

[1]A. Actor, Rev. Mod. Phys. **51**, 461 (1979).
[2]E. Witten, Phys. Rev. Lett. **38**, 121 (1977).
[3]J. Cerveró, L. Jacobs, and C. Nohl, Phys. Lett. B **69**, 351 (1977).
[4]C. Rebbi, Phys. Rev. D **17**, 483 (1978).
[5]B. Schechter, Phys. Rev. D **16**, 3015 (1977).
[6]J. Cerveró, J. Math. Phys. **23**, 1466 (1982).
[7]S. Hawking and G. Ellis, "The large scale structure of space-time," Cambridge Monogr. Math. Phys. **1973**, 118.
[8]V. de Alfaro, S. Fubini, and G. Furlan, Phys. Lett. B **72**, 203 (1976).
[9]A. Actor, Dortmund preprint.
[10]T. Wu and C. Yang, Phys. Rev. D **12**, 3843 (1975).
[11]T. Wu and C. Yang, Phys. Rev. D **13**, 3233 (1976).
[12]J. Hsu and E. Mac, J. Math. Phys. **18**, 100 (1977).
[13]W. Bernrenter, Phys. Rev. D **16**, 3609 (1977).

On total noncommutativity in quantum mechanics

Pekka J. Lahti
Department of Physical Sciences, University of Turku, 20500 Turku, Finland
Kari Ylinen[a)]
Department of Mathematics, University of California, Riverside, California 92521-0135

(Received 5 March 1987; accepted for publication 1 July 1987)

It is shown within the Hilbert space formulation of quantum mechanics that the total noncommutativity of any two physical quantities is necessary for their satisfying the uncertainty relation or for their being complementary. The importance of these results is illustrated with the canonically conjugate position and momentum of a free particle and of a particle closed in a box.

I. INTRODUCTION

The aim of this paper is to contribute to the problem of commutativity in quantum mechanics. It will be shown that any two physical quantities which either satisfy the uncertainty relation or which are complementary are also totally noncommutative. These results will be illustrated by the important examples of canonically conjugate position and momentum observables of a free particle and of a particle closed in a box.

Throughout this paper the standard Hilbert space formulation of quantum mechanics will be applied. To fix the notations and terminology we shall briefly recall its basic ingredients here.

In the Hilbert space formulation of quantum mechanics the description of a physical system is based on a (complex, separable, generally infinite-dimensional) Hilbert space $\mathbf{H}$, with the inner product $\langle \cdot | \cdot \rangle$. We let $\mathbf{L}(\mathbf{H})$ denote the set of bounded linear operators on $\mathbf{H}$, and $\mathbf{P}(\mathbf{H})$ the subset of $\mathbf{L}(\mathbf{H})$ consisting of the (orthogonal) projections. Any physical quantity of the system is represented as (and identified with) a self-adjoint operator A in $\mathbf{H}$. The spectral measure of A is denoted by $E^A : \mathbf{B}(\mathbb{R}) \to \mathbf{L}(\mathbf{H})$, where $\mathbf{B}(\mathbb{R})$ is the Borel σ-algebra of the real line $\mathbb{R}$ [and $\mathbf{B}(\mathbb{R}^2)$ that of $\mathbb{R}^2$]. Any state of the system is represented as (and identified with) an element T of the set $\mathbf{T}_s(\mathbf{H})_1^+$ of positive normalized trace class operators on $\mathbf{H}$. In this representation, the pure states of the system appear as the (normalized) one-dimensional projection operators $|\varphi\rangle\langle\varphi|$ on $\mathbf{H}$, $\varphi \in \mathbf{H}$, so that they may be identified, modulo a phase factor, with the unit vectors φ of $\mathbf{H}$. Here $\mathbf{H}_1$ denotes the set of unit vectors of $\mathbf{H}$. The probability measure E_T^A [$E_T^A(X) := \mathrm{tr}(TE^A(X))$, $X \in \mathbf{B}(\mathbb{R})$] defined by a physical quantity A and a state T is interpreted as the probability distribution of the values of the quantity A in the state T. In particular, the variance

$$\int_{\mathbb{R}} \left(i - \int_{\mathbb{R}} i \, dE_T^A \right)^2 dE_T^A$$

(where i: $\mathbb{R} \to \mathbb{R}$ is the identity map) of the probability measure E_T^A will be denoted by $\mathrm{Var}(A,T)$, and it is interpreted as the variance of A in the state T. (For further details of this formulation of quantum mechanics we refer to Beltrametti and Cassinelli.[1]).

In studying the commutativity of two physical quantities A and B their *commutativity domain* $\mathrm{com}(A, B) := \{\varphi \in \mathbf{H} | E^A(X)E^B(Y)\varphi = E^B(Y)E^A(X)\varphi$ for all $X,Y \in \mathbf{B}(\mathbb{R})\}$ has turned out to be highly useful (see, e.g., Hardegree,[2] Pulmannova,[3] Pulmannova and Dvurecenskij,[4] Lahti,[5] Ylinen,[6] and Busch and Lahti[7]). Here $\mathrm{com}(A,B)$ is a closed subspace of $\mathbf{H}$, which is invariant under the spectral projections of A and B, i.e., $E^A(X)(\mathrm{com}(A,B)) \subset \mathrm{com}(A,B)$ and $E^B(Y)(\mathrm{com}(A,B)) \subset \mathrm{com}(A,B)$ for all $X,Y \in \mathbf{B}(\mathbb{R})$. Clearly, A and B commute (in the sense that all their spectral projections commute) if and only if $\mathrm{com}(A,B) = \mathbf{H}$.

The notions of commutativity and commutativity domain admit both probabilistic and operational characterizations. On the one hand, $\mathrm{com}(A,B)$ consists exactly of those (vector) states $\varphi \in \mathbf{H}_1$ for which the map $X \times Y \mapsto \langle \varphi | E^A(X) \wedge E^B(Y)\varphi \rangle$, $X,Y \in \mathbf{B}(\mathbb{R})$, extends to a probability measure $\mathbf{B}(\mathbb{R}^2) \to [0,1]$, $Z \mapsto E^{A,B}_{|\varphi\rangle\langle\varphi|}(Z)$.[4,6,8] On the other hand, $\varphi \in \mathrm{com}(A,B)$ if and only if $\Phi_X^A \cdot \Phi_Y^B \times |\varphi\rangle\langle\varphi| = \Phi_Y^B \cdot \Phi_X^A |\varphi\rangle\langle\varphi|$ for all $X,Y \in \mathbf{B}(\mathbb{R})$, i.e., the von Neumann–Lüders measurements Φ_X^A and Φ_Y^B of A and B associated with the value sets X and Y commute in the state $|\varphi\rangle\langle\varphi|$.[9] [Here, e.g., Φ_X^A denotes the von Neumann–Lüders operation $T \mapsto \Phi_X^A T := E^A(X)TE^A(X)$, $T \in \mathbf{T}_s(\mathbf{H})_1^+$]. These results can readily be generalized to arbitrary states $T \in \mathbf{T}_s(\mathbf{H})_1^+$, but to avoid some technical details we omit the more general formulations here.

Physical quantitites A and B are *totally noncommutative* if $\mathrm{com}(A,B) = \{0\}$, i.e., if 0 is the only vector with respect to which the spectral projections of A and B commute. From the probabilistic point of view this is to say that A and B have a joint probability $E^{A,B}_{|\varphi\rangle\langle\varphi|}$ in no state $|\varphi\rangle\langle\varphi|$, $\varphi \in \mathbf{H}_1$. No probabilistic predictions on their joint values can then be done. The operational content of this result appears most directly in the fact that in any state $|\varphi\rangle\langle\varphi|$, $\varphi \in \mathbf{H}_1$, some of the von Neuman–Lüders sequential measurements $\Phi_X^A \cdot \Phi_Y^B$ and $\Phi_Y^B \cdot \Phi_X^A$ are distinguishable, i.e., independently of the state of the system the order in which A and B are measured in sequence is, in general, relevant to the measuring result.

In the next two sections we shall show that any two physical quantities that either satisfy the uncertainty relation or are complementary are also totally noncommutative. In Sec. IV we shall briefly demonstrate the relevance of these

[a)] Permanent address: Department of Mathematics, University of Turku, 20500 Turku, Finland.

 0022-2488/87/112614-04$02.50

results for position and momentum observables of a "free particle" as well as of a "particle closed in a box."

II. UNCERTAINTY RELATION AND TOTAL NONCOMMUTATIVITY

Let A and B be any two physical quantities. We shall now study the product $\mathrm{Var}(A,T)\cdot\mathrm{Var}(B,T)$ of their variances in a state $T\in\mathbf{T}_s(\mathbf{H})_1^+$. We denote $h(A,B) = \inf\{\mathrm{Var}(A,T)\cdot\mathrm{Var}(B,T)\,|\,T\in\mathbf{T}_s(\mathbf{H})_1^+\}$ so that $h(A,B) = 0$ or $h(A,B) > 0$. If $h(A,B) > 0$ we say that A and B *satisfy the uncertainty relation*. In that case the "uncertainty product" $\mathrm{Var}(A,T)\cdot\mathrm{Var}(B,T)$ of A and B has a state independent positive lower bound.

Lemma 2.1: If $X\subset\mathbb{R}$ is a Borel set contained in an interval $[a,b]$, and $\varphi\in E^A(X)(\mathbf{H})$, $\|\varphi\| = 1$, then $\mathrm{Var}(A,|\varphi\rangle\langle\varphi|)\leqslant(b-a)^2$.

Proof: The probability measure $E^A_{|\varphi\rangle\langle\varphi|}$ satisfies

$$\begin{aligned} E^A_{|\varphi\rangle\langle\varphi|}(Y) &= \langle\varphi\,|E^A(Y)\varphi\rangle \\ &= \langle E^A(X)\varphi\,|E^A(Y)E^A(X)\varphi\rangle \\ &= \langle E^A(X)\varphi\,|E^A(Y\cap X)\varphi\rangle \end{aligned}$$

for all $Y\in\mathbf{B}(\mathbb{R})$, so that it is concentrated on $[a,b]$. Thus $\int i\, dE^A_{|\varphi\rangle\langle\varphi|} = c\in[a,b]$, and so

$$\mathrm{Var}(A,|\varphi\rangle\langle\varphi|) = \int_{[a,b]}(i-c)^2\, dE^A_{|\varphi\rangle\langle\varphi|}\leqslant(b-a)^2.$$

■

Proposition 2.2: If $\mathrm{com}(A,B)\neq\{0\}$, then $h(A,B) = 0$.

Proof: Suppose $\varphi\in\mathrm{com}(A,B)$, $\varphi\neq0$. Given $\epsilon>0$, let $(X_i)_{i=1}^{\infty}$ and $(Y_j)_{j=1}^{\infty}$ be two partitions of $\mathbb{R}$ into unions of disjoint half-open intervals of length $\delta = \min\{\epsilon,1\}$. Since multiplication in $\mathbf{L}(\mathbf{H})$ is separately continuous in the strong operator topology, we get

$$\begin{aligned} \varphi &= \Big[\sum_{i=1}^{\infty}E^A(X_i)\Big]\Big[\sum_{j=1}^{\infty}E^B(Y_j)\Big]\varphi \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}E^A(X_i)E^B(Y_j)\varphi, \end{aligned}$$

so that for some i and j, $\psi_0 = E^A(X_i)E^B(Y_j)\varphi\neq0$. Denote $\psi = \|\psi_0\|^{-1}\psi_0$. Then $\psi\in E^A(X_i)(\mathbf{H})$, and since $E^A(X_i)E^B(Y_j)\varphi = E^B(Y_j)E^A(X_i)\varphi$, we also have $\psi\in E^B(Y_j)(\mathbf{H})$. Applying Lemma 2.1 we thus obtain $\mathrm{Var}(A,|\psi\rangle\langle\psi|)\cdot\mathrm{Var}(B,|\psi\rangle\langle\psi|)\leqslant\delta^4\leqslant\epsilon$. ■

Corollary 2.3: If A and B satisfy the uncertainty relation, then they are totally noncommutative.

Remark 2.4: The result 2.3 was originally obtained within the so-called quantum logic frame by Pulmannova and Dvurecenskij.[4] The result is formulated here within the Hilbert space frame of quantum mechanics so that the proof is now more direct.

III. COMPLEMENTARITY AND TOTAL NONCOMMUTATIVITY

Experimental arrangements which admit unambiguous operational definitions of complementary physical quantities are mutually exclusive. This old intuitive idea of Bohr and Pauli has systematically been developed by Lahti[5,10,11] and by Beltrametti and Cassinelli.[1] The notion of complementary physical quantities is based on the mutual exclusiveness of any two instruments (i.e., operation valued measures) which uniquely define these quantities. Here we do not need the notion of instrument. Thus we shall adopt here a formally equivalent though less informative definition of the notion of complementarity of physical quantities: Two physical quantities A and B are *complementary* if $E^A(X)\wedge E^B(Y) = 0$ for all bounded $X,Y\in\mathbf{B}(\mathbb{R})$ for which neither $E^A(X)$ nor $E^B(Y)$ equals I.

We say that a physical quantity A is *constant* if 0 and I are the only spectral projections A has. Clearly, if A is constant it commutes and is complementary with any other quantity B. The converse result is an immediate consequence of Proposition 3.2. Before proving it we shall state a lemma which characterizes nonconstant physical quantities through some simple properties of their spectral measures.

Lemma 3.1: For a spectral measure E: $\mathbf{B}(\mathbb{R})\to\mathbf{L}(\mathbf{H})$ the following conditions are equivalent: (i) $E(\mathbf{B}(\mathbb{R}))\neq\{0,1\}$; (ii) supp (E) $[:=\cap(X\in\mathbf{B}(\mathbb{R})|X$ is closed and $E(X) = I)]$ contains more than one point; (iii) there is a partition $(X_i)_{i=1}^{\infty}$ of $\mathbb{R}$ into a union of disjoint bounded intervals such that $E(X_i)\neq I$ for all $i = 1,2,\ldots$.

Proof: Clearly (i)$\Rightarrow$(ii), for if $\mathrm{supp}(E) = \{a\}$, then $E(X) = I$ if $a\in X$, and $E(X) = 0$ otherwise. If (ii) holds true, then we get (iii) by choosing any partition (X_i) of $\mathbb{R}$ such that two distinct points of $\mathrm{supp}(E)$ are interior points of distinct intervals X_i. The implication (iii)$\Rightarrow$(i) is even more immediate. ■

Proposition 3.2: Let A and B be two nonconstant complementary physical quantities. Then $\mathrm{com}(A,B) = \{0\}$.

Proof: Let us assume that $\mathbf{K} = \mathrm{com}(A,B)\neq\{0\}$. Since $\mathbf{K}$ is a closed subspace which is invariant under the spectral projections of A and B, we obtain two spectral measures $\widetilde{E}^A$: $\mathbf{B}(\mathbb{R})\to\mathbf{L}(\mathbf{K})$ and $\widetilde{E}^B$: $\mathbf{B}(\mathbb{R})\to\mathbf{L}(\mathbf{K})$ by setting $\widetilde{E}^A(X) = E^A(X)|_{\mathbf{K}}:\mathbf{K}\to\mathbf{K}$ for $X\in\mathbf{B}(\mathbb{R})$, and defining $\widetilde{E}^B$ similarly in terms of E^B.[6] Choose partitions (X_i) and (Y_j) for E^A and E^B as in Lemma 3.1. Then $I_{\mathbf{K}} = \Sigma_{i=1}^{\infty}E^A(X_i) = \Sigma_{j=1}^{\infty}E^B(Y_j)$, the projections $\widetilde{E}^A(X_i)$ and $\widetilde{E}^B(Y_j)$ commute, and $[\widetilde{E}^A(X_i)\wedge\widetilde{E}^B(Y_j)]\varphi = [E^A(X_i)\wedge E^B(Y_j)]\varphi$ for all $\varphi\in\mathbf{K}$.[6] It follows that $0\neq\widetilde{E}^A(X_i)\widetilde{E}^B(Y_j) = \widetilde{E}^A(X_i)\wedge\widetilde{E}^B(Y_j) = 0$ for some i,j. This contradiction proves that $\mathbf{K} = \{0\}$. ■

Remark 3.3: The fact that nonconstant complementary quantities are necessarily noncommutative was obtained by Lahti[10] within a more general quantum logic frame. The stronger result 3.2 was claimed without proof by Pulmannova and Dvurecenskij[4] also in a more general context. Here the result is obtained within the Hilbert space quantum mechanics applying the methods developed by Ylinen.[6]

IV. TWO EXAMPLES

In this section we illustrate the above results with two examples. The first essentially refers to the canonically conjugate position q and momentum p of a free particle (in $\mathbb{R}$), and the second to the same observables of a particle closed in a (one-dimensional) box (of unit length).

Example 4.1: Let $\mathbf{H}$ be the Lebesgue space $L^2(\mathbb{R},\mathbf{B}(\mathbb{R}), dx,\mathbb{C})$, and let F denote the Fourier–Plancherel operator on

$\mathbf{H}$. We define two spectral measures $E^q: \mathbf{B}(\mathbb{R}) \to \mathbf{L}(\mathbf{H})$ and $E^p: \mathbf{B}(\mathbb{R}) \to \mathbf{L}(\mathbf{H})$ by the formulas $E^q(X)f = \chi_X f$ for all $X \in \mathbf{B}(\mathbb{R})$, $f \in \mathbf{H}$, and $E^p(Y) = F^{-1}E^q(Y)F$ for all $Y \in \mathbf{B}(\mathbb{R})$, and denote the corresponding self-adjoint operators in $\mathbf{H}$ by q and p. It is well known that q is the multiplication operator, $(qf)(x) = xf(x)$, $x \in \mathbb{R}$, with the domain

$$\mathbf{D}(q) = \left\{ f \in \mathbf{H} \,\middle|\, \int_{\mathbb{R}} i^2 \, dE^q_{|f\rangle\langle f|} < \infty \right\},$$

where $E^q_{|f\rangle\langle f|}(X) = \langle f | E^q(X) f \rangle$ (inner product in $\mathbf{H}$). Similarly, it is known that the domain of p is $\mathbf{D}(p) = \{ f \in \mathbf{H} \,|\, f$ is absolutely continuous on every compact interval and $f' \in \mathbf{H}\}$, and $(pf)(x) = -i(d/dx)f(x)$ (a.e.) for $f \in \mathbf{D}(p)$ (see, e.g., Stone,[12] p. 441). For example, if f is in $C_c^\infty(\mathbb{R})$, the space of infinitely differentiable functions with compact support, a straightforward calculation shows that $(qp - pq)f = if$, i.e., the formula

$$qp - pq = iI \tag{1}$$

holds on the dense subspace $C_c^\infty(\mathbb{R})$ of $\mathbf{H}$. Moreover, as $\mathrm{Var}(q,f) \cdot \mathrm{Var}(p,f) = \mathrm{Var}|f|^2 \cdot \mathrm{Var}|\hat{f}|^2 \geqslant \frac{1}{4}$ for all $f \in \mathbf{H}$, and the lower bound $\frac{1}{4}$ is reached with any Gaussian probability density function (see, e.g., Cowling and Price,[13] p. 152, but compare also Busch and Lahti[14] for historical remarks) we have the well-known formula

$$h(q,p) = \tfrac{1}{2}. \tag{2}$$

(Here $\hat{f}$ is the Fourier transform of f.) The property

$$E^q(X) \wedge E^p(Y) = 0 \quad \text{for all bounded } X, Y \in \mathbf{B}(\mathbb{R}) \tag{3}$$

is also well known (see, e.g., Lenard[15], but compare also Busch and Lahti[14]). According to Corollary 2.3 and Theorem 3.2 both (2) and (3) imply that q and p are also totally noncommutative, i.e.,

$$\mathrm{com}(q,p) = \{0\}. \tag{4}$$

(Compare Hardegree[2], p. 505.) As is well known, the above operators q and p represent the canonically conjugate position and momentum of a free particle (in $\mathbb{R}$).

Example 4.2: Let m be the normalized Haar measure of the compact Abelian group $\mathbb{T} = \{z \in \mathbb{C} \,|\, |z| = 1\}$ (i.e., m is the Lebesgue measure on the Borel σ-algebra of $[0,1)$ transferred by the map $t \mapsto e^{i2\pi t}$), and $\mathbf{H} = L^2(m) = L^2(\mathbb{T}, \mathbf{B}(\mathbb{T}), m, \mathbb{C})$. Here $\mathbb{T}$ can be identified with the quotient group $\mathbb{R}/\mathbb{Z}$, and then the dual group of $\mathbb{T}$ is identified with $2\pi\mathbb{Z}$. For $f \in \mathbf{H}$ define as usual $\hat{f}: 2\pi\mathbb{Z} \to \mathbb{C}$ by $\hat{f}(2\pi n) = \int_{\mathbb{T}} f(z) \bar{z}^n \, dm(z)$. It is well known that we get an isometric isomorphism $F: L^2(m) \to l^2(2\pi\mathbb{Z})$ by $F(f) = \hat{f}$; this is the Fourier–Plancherel transformation. We again define two spectral measures $Q: \mathbf{B}(\mathbb{T}) \to \mathbf{L}(\mathbf{H})$ and $P: \mathbf{B}(2\pi\mathbb{Z}) \to \mathbf{L}(\mathbf{H})$ by the formulas $Q(X)f = \chi_X f$, $X \in \mathbf{B}(\mathbb{T})$, and $P(Y)f = F^{-1}\chi_Y Ff$, $Y \in \mathbf{B}(2\pi\mathbb{Z})$. First we note that

$$Q(X) \wedge P(Y) = 0, \tag{5}$$

whenever $X \in \mathbf{B}(\mathbb{T})$ is a set with $m(X) < 1$ and $Y \subset 2\pi\mathbb{Z}$ is a finite set. Indeed, if $f \in P(Y)(\mathbf{H})$ for a finite set $Y \subset 2\pi\mathbb{Z}$, then f is a trigonometric polynomial which implies that $f \equiv 0$ or f vanishes only at finitely many points, so that if $f \in Q(X)(\mathbf{H})$ where $m(X) < 1$, then $f = 0$. Now let $\theta: [0,1) \to \mathbb{T}$ be the map $t \mapsto e^{i2\pi t}$ and define $Q_0(X) = Q(\theta(X \cap [0,1)))$, $X \in \mathbf{B}(\mathbb{R})$. For $Y \in \mathbf{B}(\mathbb{R})$, define $P_0(Y) = P(Y \cap 2\pi\mathbb{Z})$. Then Q_0 and P_0 are spectral measures on $\mathbf{B}(\mathbb{R})$, and from (1) it follows at once that the corresponding physical quantities $q = \int_{\mathbb{R}} x \, dQ_0(x)$ and $p = \int_{\mathbb{R}} x \, dP_0(x)$ are complementary, i.e.,

$$Q_0(X) \wedge P_0(Y) = 0 \quad \text{for all bounded } X, Y \in \mathbf{B}(\mathbb{R}) \text{ for which } Q_0(X) \neq I \neq P_0(Y). \tag{6}$$

Thus by Theorem 3.2 they are totally noncommutative, i.e.,

$$\mathrm{com}(q,p) = \{0\}. \tag{7}$$

(In Ylinen[16,17] this is proved in a different way.) Since the spectrum of p is $2\pi\mathbb{Z}$, and thus discrete, q and p do not satisfy the uncertainty relation; we have

$$h(q,p) = 0. \tag{8}$$

Naturally, $L^2([0,1])$ can be identified with $L^2([0,1])$, so that θ defines an isometric isomorphism $U: \mathbf{H} \to L^2([0,1])$. Thus Q_0 and P_0 can be regarded as maps into $\mathbf{L}(L^2([0,1]))$. With this interpretation, q and p are simply the canonically conjugate position and momentum observables of a particle closed in a one-dimensional box of unit length. In fact, in Ylinen[16,18] it is shown that $\mathbf{D}(q) = L^2([0,1])$, $(qf)(x) = xf(x)$, $\mathbf{D}(p) = \{ f \in L^2([0,1]) \,|\, f$ is absolutely continuous, $f(0) = f(1)$, and $f' \in L^2([0,1])\}$, and $(pf)(x) = -i(d/dx)f(x)$ (a.e.). From this one then gets

$$qp - pq = iI \tag{9}$$

on a dense subspace of $L^2([0,1])$ (e.g., on $\{f: [0,1] \to \mathbb{C} \,|\, f$ is infinitely differentiable and $\mathrm{supp}(f) \subset (0,1)\}$).

Remark 4.3: The two (q,p) pairs of examples 4.1 and 4.2 manifestly differ in the fact that the first pair satisfies the uncertainty relation while the second does not. In spite of that, the two examples share a common general structure: each one is a concrete and physically important realization of an abstract Weyl pair on a Hilbert space $\mathbf{H}$, based on a locally compact Abelian group G and its dual group $\hat{G}$. In the first case G is the group of translations of $\mathbb{R}$ so that G and $\hat{G}$ can be identified with the additive group $\mathbb{R}$. The relevant Hilbert space $\mathbf{H}$ is then $L^2(\mathbb{R}, \mathbf{B}(\mathbb{R}), dx, \mathbb{C})$. In the second example G is the group of translations on $[0,1)$ (mod 1). As noted there, G may be identified with the torus group $\mathbb{T}$, and $\hat{G}$ with $2\pi\mathbb{Z}$, and in this case the Hilbert space is $L^2(\mathbb{T}, \mathbf{B}(\mathbb{T}), m, \mathbb{C})$. For details of obtaining the pair (q,p) from the duality between $\mathbb{T}$ and $2\pi\mathbb{Z}$ we refer to Ylinen[16,18] whereas the first case is well known (see, e.g., von Neumann[19]).

[1]E. Beltramentti and G. Cassinelli, *The Logic of Quantum Mechanics* (Addison–Wesley, Reading, MA, 1981).
[2]G. Hardegree, "Relative compatibility in conventional quantum mechanics," Found. Phys. 7, 495 (1977).
[3]S. Pulmannova, "Relative compatibility and joint distributions of observables," Found. Phys. **10**, 641 (1980).
[4]S. Pulmannova and A. Dvurecenskij, "Uncertainty principle and joint dis-

tribution of observables," Ann. Inst. H. Poincaré A **42**, 253 (1985).
[5]P. J. Lahti, "Uncertainty, complementarity, and commutativity," in *Recent Developments in Quantum Logic*, edited by P. Mittelstaedt and E.-W. Stachow (BI Wissenschaftsverlag, Mannheim, 1985).
[6]K. Ylinen, "On a theorem of Gudder on joint distributions of observables," in *Symposium on the Foundations of Modern Physics*, edited by P. Lahti and P. Mittelstaedt (World Scientific, Singapore, 1985), pp. 691–694.
[7]P. Busch and P. Lahti, "To what extent do position and momentum commute?" Phys. Lett. A **115**, 259 (1986).
[8]S. P. Gudder, "Joint distributions of observables," J. Math. Mech. **18**, 325 (1968).
[9]P. J. Lahti, "Commutativity, probabilistic complementarity, and maximal information in quantum mechanics," Report Series Turku-FTL-R104, University of Turku, 1986.
[10]P. J. Lahti, "Uncertainty and complementarity in axiomatic quantum mechanics," Int. J. Theor. Phys. **19**, 789 (1980).
[11]P. J. Lahti, "Hilbertian quantum theory as the theory of complementarity," Int. J. Theor. Phys. **22**, 911 (1983).
[12]M. H. Stone, *Linear Transformations in Hilbert Space and Their Applications to Analysis* (American Mathematical Society Colloquium Publications, New York, 1932), Vol. XV.
[13]M. Cowling and J. Price, "Bandwith versus time concentration: The Heisenberg–Pauli–Weyl inequality," SIAM J. Math. Anal. **15**, 151 (1984).
[14]P. Busch, and P. Lahti, "A note on quantum theory, complementarity, and uncertainty," Philos. Sci. **52**, 64 (1985).
[15]A. Lenard, "The numerical range of a pair of projections," J. Funct. Anal. **10**, 410 (1972).
[16]K. Ylinen, "Commutation properties of observables as spectral measures on locally compact abelian groups," University of California, Riverside, Department of Mathematics, Technical Report, March 1987.
[17]K. Ylinen, "Position and momentum as spectral measures on locally compact abelian groups," in *Symposium on the Foundations of Modern Physics*, edited by P. Lahti and P. Mittelstaedt (World Scientific, Singapore, to appear).
[18]K. Ylinen, "Commuting functions of the position and momentum observables on locally compact abelian groups," J. Math. Anal. Appl. (to appear).
[19]J. von Neumann, "Die Eindeutigkeit der Schrödingerschen Operatoren," Math. Ann. **104**, 570 (1931).

The connection of two-particle relativistic quantum mechanics with the Bethe–Salpeter equation

H. Sazdjian
Division de Physique Théorique,[a] *Institut de Physique Nucléaire, Université Paris XI, 91406 Orsay Cedex, France*

(Received 14 January 1987; accepted for publication 3 June 1987)

The formal equivalence between the wave equations of two-particle relativistic quantum mechanics, based on the manifestly covariant Hamiltonian formalism with constraints, and the Bethe–Salpeter equation are shown. This is achieved by algebraically transforming the latter so as to separate it into two independent equations that match the equations of Hamiltonian relativistic quantum mechanics. The first equation determines the relative time evolution of the system, while the second one yields a three-dimensional eigenvalue equation. A connection is thus established between the Bethe–Salpeter wave function and its kernel on the one hand and the quantum mechanical wave function and interaction potential on the other. For the sector of solutions of the Bethe–Salpeter equation having nonrelativistic limits, this relationship can be evaluated in perturbation theory. A generalized form of the instantaneous approximation that simplifies the various expressions involved in the above relations is also devised. It also permits the evaluation of the normalization condition of the quantum mechanical wave function as a three-dimensional integral.

I. INTRODUCTION

The interpretation of the Bethe–Salpeter equation[1–3] as a relativistic wave equation for interacting particle systems has met difficulties in the framework of relativistic quantum mechanics. The reason for this is the dynamical role played by the relative time variables, leading, in general, to relative energy excitations in the spectrum of the bound states.[4,5] These new excited states, the so-called "abnormal" solutions, do not have the usual nonrelativistic behavior of massive states when the velocity of light c goes to infinity. Instead, in that limit, they disappear from the spectrum.

On the other hand, in the Hamiltonian description of particle systems the time component of each coordinate four-vector is not assigned a dynamical role, but is considered as a parameter. Similarly the energy component of each momentum four-vector is expressed in terms of the other independent variables of the system. Thus for a system of particles interacting at a distance (i.e., by means of potentials) the number of degrees of freedom is equal to the sum of the degrees of freedom of each particle. Because this way of describing particle systems is closer to physical experience that Hamiltonian relativistic quantum mechanisms of a finite number of interacting particles[6] is an appropriate tool for the study of systems of particles, where radiation and inelastic effects have been neglected or treated approximately. It is also for this reason that the knowledge of the Hamiltonian content of the Bethe–Salpeter equation is of major interest.

The attempts to construct a relativistic mechanics of interacting particle systems have met, for a long time, conceptual and technical difficulties.[7] It is only in recent years that a definite progress was achieved with the aid of the manifestly covariant formalism with constraints.[8–10] The point which was still missing in the latter approach concerned its relation with field theory and, in particular, with the Bethe–Salpeter equation, where the wave function has a precise definition in terms of the local fields of the theory and where the interaction kernel is determined by means of the interaction Lagrangian.

It is the purpose of the present paper to exhibit the connection, in the two-particle case, of the manifestly covariant wave equations of Hamiltonian relativistic quantum mechanics with the Bethe–Salpeter equation. We shall show that the former can actually be derived from the latter and contain at least the sector of solutions which have Galilei invariant nonrelativistic limits (the so-called "normal" solutions).

This connection is established by transforming the Bethe–Salpeter (BS) equation by algebraic manipulations so as to separate it into two independent equations that have the same structure as the wave equations of relativistic quantum mechanics. The first equation determines the relative time evolution of the system, while the second one yields a three-dimensional eigenvalue equation. The BS wave function and the BS kernel on the one hand, the quantum mechanical wave function and the corresponding potential on the other, are connected, respectively, to each other by definite relations. These involve integral operators containing the BS kernel and could, in principle, be evaluated in perturbation theory for the sector of normal solutions of the BS equation.[11]

The question concerning the validity of the above connection also for the sectors of abnormal solutions of the BS equation is not examined in this paper, but several possibilities are invoked in the concluding remarks (Sec. VII) for future investigations.

Another method of approach, the "quasipotential" approach, was also developed in the past to reduce the Bethe-Salpeter equation to a three-dimensional one.[12–18] Here one derives a reduced form of the BS equation by starting from two-times four-point Green's functions.[12] Thus, from the start, one eliminates the sectors of relative energy excitations. This equation is also usually transformed into a Lipp-

[a] Laboratory associated with C.N.R.S.

mann–Schwinger-type equation.

Our approach differs from the quasipotential approach in that it is formulated in configuration space and uses wave functions as basic ingredients, rather than scattering amplitudes. For this reason it is more amenable for quantum mechanical problems. The quasipotential approach seems to be more suitable for scattering problems, where the basic ingredients are scattering amplitudes and where the direct connection with the local fields of the underlying field theory becomes less transparent. Wave functions are defined in a more indirect way.

Because their starting point is the Bethe–Salpeter equation as far as the sector of normal solutions is concerned, and because the transformations introduced are of algebraic nature, we think that the quasipotential approach[12] and our approach are equivalent, when considered in their general and complete forms, that is, before any specific approximation is utilized. The advantages of each approach depend, however, as we outlined above, on the framework or the type of problem considered.

In order to be able to extract the main physical information from the relations we established between the quantum mechanical and the BS equations, we devise an approximation, which we call "the relativistic instantaneous approximation." It is a generalization, in some sense, of the old "instantaneous approximation,"[11,19] and has four main advantages: first, it is manifestly covariant and does not destroy the relativistic invariance of the system; second, it does not necessitate any truncation of the interaction kernel; third, it permits a partial summation of the iteration series involved in the exact relations; and, fourth, it automatically selects the sector of normal solutions of the BS equation. This approximation leads to a considerable simplification of the expressions of the relations and will be widely utilized throughout this paper. In particular, in this approximation, and when the radiative corrections of the external particle propagators are neglected, the covariant relative time dependence of the BS wave function becomes completely determined in an explicit and kinematic way, irrespective of the form of the BS kernel.

The relativistic instantaneous approximation is also close in spirit to the method introduced by Blankenbecler and Sugar[13] in the framework of the quasipotential approach to reduce the BS equation to a three-dimensional one and to simplify its treatment. Here one starts with the BS equation written as an integral equation for off-mass-shell scattering amplitudes. One then subtracts from and adds to the kernel K of the equation another kernel $\overline{K}$, appropriately chosen, such that $\overline{K}$, rather than K, serves as the starting point for the physical calculations and reduces at the same time the integral equation to a three-dimensional one. The rest of the kernel, $K-\overline{K}$, is then treated as a perturbation in an iteration series. There are, of course, many choices for the kernel $\overline{K}$,[13–15,18] but these, in general, take into account the physical properties of the system at low energies, such as the unitarity condition and the nonrelativistic limit.

The relationship between the BS wave function and the quantum mechanical wave function also permits one to obtain the normalization condition of the latter from that of the former.[3,20,21] It is obtained, in its relativistic instantaneous approximation, as a three-dimensional integral.

The plan of the paper is as follows. In Sec. II we consider the case of two spin-0 boson systems. (The results concerning two equal mass spin-0 bosons were already presented in Ref. 22) In Sec. III, we introduce the relativistic instantaneous approximation and apply it to the general results obtained in Sec. II. The case of spin-$\frac{1}{2}$ fermion–antifermion systems is treated in Sec. IV, while systems composed of one spin-$\frac{1}{2}$ fermion and one spin-0 boson are considered in Sec. V. The normalization condition of the quantum mechanical wave function is presented in Sec. VI. The Conclusion follows in Sec. VII.

Throughout this paper we always consider bound states, or two-particle scattering states, which have a real and positive total mass squared, $p^2>0$, and positive total mass, $(p^2)^{1/2}>0$.

II. TWO SPIN-0 BOSON SYSTEMS

In the manifestly covariant relativistic quantum mechanics the wave function of the two spin-0 particle system satisfies two independent wave equations,[10] each of them being a generalization of the Klein–Gordon equation for particle 1 or 2,

$$H_a\Psi(x_1,x_2)\equiv(p_a^2-m_a^2-V)\Psi(x_1,x_2)=0\quad(a=1,2),\tag{2.1}$$

where V is a Poincaré invariant interaction potential.[23] These two wave equations must, however, be compatible among themselves and one therefore gets the compatibility (or integrability) condition

$$[H_1,H_2]\Psi=0.\tag{2.2}$$

This equation is a constraint on the potential. By demanding that it be satisfied in the strong sense one gets

$$[p_1^2-p_2^2,V]=0.\tag{2.3}$$

The Poincaré invariant general (local) solution of this equation is

$$V=V(x^{T2},x^T\cdot v,p^2,v^2,p\cdot v),\tag{2.4}$$

where we use the following notation:

$$\begin{aligned}&p=p_1+p_2,\quad v=\tfrac{1}{2}(p_1-p_2),\\&X=\tfrac{1}{2}(x_1+x_2),\quad x=x_1-x_2,\\&x_\mu^T=x_\mu-(\hat p\cdot x)\hat p_\mu,\quad \hat p_\mu=p_\mu/(p^2)^{1/2},\\&p^2>0,\quad x^{T^2}=x^2-(\hat p\cdot x)^2.\end{aligned}\tag{2.5}$$

For any vector Y we shall define its "transverse" and "longitudinal" parts with respect to the total momentum p by

$$Y_\mu^T=Y_\mu-(\hat p\cdot Y)\hat p_\mu,\quad Y_\mu^L=(\hat p\cdot Y)\hat p_\mu,\quad Y_L=(\hat p\cdot Y).\tag{2.6}$$

By subtracting the two equations (2.1) from each other we get an equation that determines the covariant relative time evolution of the system:

$$(p_1^2-p_2^2)\Psi=2p\cdot v\Psi=(m_1^2-m_2^2)\Psi.\tag{2.7}$$

The solution of this, for eigenfunctions of p_μ, is

$$\Psi(x_1,x_2)=e^{-ip\cdot X}e^{-i(m_1^2-m_2^2)p\cdot x/(2p^2)}\psi(x^T),\tag{2.8}$$

and $\psi(x^T)$ defines a three-dimensional "internal" wave function. Taking the sum of the two equations (2.1), one gets the "eigenvalue" equation

$$[\tfrac{1}{4}p^2 - \tfrac{1}{2}(m_1^2 + m_2^2) + (1/4p^2)(m_1^2 - m_2^2)^2 + v^{T^2} - V]\psi(x^T) = 0, \qquad (2.9)$$

which is a three-dimensional Schrödinger-type equation. If the potential V is chosen to be of order c^2 (c is the velocity of light), then, in the nonrelativistic limit, Eq. (2.9) reduces to a Galilei invariant Schrödinger equation.[24]

The wave equations (2.1) can also be generalized so as to include nonlocal potentials (in x^T):

$$(p_a^2 - m_a^2)\Psi(X,x) = \int V(x^T,x'^T;p)\Psi(X,x_L,x'^T)d^3x'^T \quad (a=1,2), \qquad (2.10)$$

where V is a Poincaré invariant function, may depend on p^2 and also act as a derivative in the relative (transverse) coordinates, and x'^T is defined as the transverse part of x' with respect to p, with $x'_L = x_L$. As a consequence of Eqs. (2.10) the wave function Ψ has the same dependence on the longitudinal variable x_L as in Eq. (2.8) and one ends up with a three-dimensional nonlocal Schrödinger-type equation.

Up to Poincaré invariant canonical transformations,[23] Eqs. (2.10) are the most general Poincaré invariant wave equations satisfying the compatibility condition (2.2). Therefore, one has to expect that any other Poincaré invariant description of a two spin-0 particle system, having a nonrelativistic Galilei invariant limit, is equivalent, by means of some nonsingular transformation, to the one obtained with Eqs. (2.10). It appears that this is actually the case for the Bethe–Salpeter equation and we intend to show in the rest of this section, the existence of such a transformation.

We first notice that the quantum mechanical wave function Ψ, (2.8), cannot be identified, except in the free case, with the BS wave function Φ. The reason is that the latter exhibits a complicated relative time dependence (in x_L) which is dictated by its spectral properties in momentum space; in sum, it has to satisfy a Deser–Gilbert–Sudarshan representation,[25] which is typical for vertex functions, whereas the relative time dependence (in x_L) of Ψ, (2.8), is rather trivial. We therefore have to expect that the quantum mechanical wave function Ψ is related to the BS wave function Φ by some nontrivial transformation, which in turn should relate the BS equation to Eqs. (2.10).[26]

To find this relationship we start from the Bethe–Salpeter equation written in its integrodifferential form:

$$(p_1^2 - m_1^2)\Phi(x_1,x_2) = -i\int d^4y_1\, d^4y_2\, d^4x_1'\, d^4x_2'\, K\left(y_1,y_2,x_1',x_2';i\frac{\partial}{\partial x_1},i\frac{\partial}{\partial x_2}\right) \times\tfrac{1}{2}\Delta_{2F}(x_2-y_2,m_2)\delta^4(x_1-y_1)\Phi(x_1',x_2'), \qquad (2.11a)$$

$$(p_2^2 - m_2^2)\Phi(x_1,x_2) = -i\int d^4y_1\, d^4y_2\, d^4x_1'\, d^4x_2'\, K\left(y_1,y_2,x_1'x_2';i\frac{\partial}{\partial x_1},i\frac{\partial}{\partial x_2}\right) \times\tfrac{1}{2}\Delta_{1F}(x_1-y_1,m_1)\delta^4(x_2-y_2)\Phi(x_1',x_2'), \qquad (2.11b)$$

where $\frac{1}{2}\Delta_F(x,m)$ is the free propagator[27]

$$\frac{1}{2}\Delta_F(x,m) = \frac{i}{(2\pi)^4}\int\frac{e^{-ik\cdot x}}{k^2-m^2+i\epsilon}d^4k = \frac{1}{2(m^2-v^{T2})^{1/2}} \times\exp[-i|x_L|(m^2-v^{T2})^{1/2}]\delta^3(x^T), \qquad (2.12)$$

m_a ($a=1,2$) representing the renormalized masses, and we have included, in the form of derivative operators acting on x_1 and x_2, all radiative corrections of the full propagators $\frac{1}{2}\Delta'_{aF}$ ($a=1,2$), in the kernel K.

To see the latter feature, we notice that we are considering nonconfining systems, in which case the pole terms in the renormalized masses of the full propagators can be factored out. In this case the full propagator $\frac{1}{2}\Delta'_F(x,m)$ satisfies the representation

$$\frac{1}{2}\Delta'_{aF}(x_a,m_a) = \frac{i}{(2\pi)^4}\int\frac{e^{-ik\cdot x_a}R_a(k)}{k^2-m_a^2+i\epsilon}d^4k \quad (a=1,2),$$

where $R_a(k)$ represents the contribution of the radiative corrections to the free propagator, other than the additive corrections included in the renormalized mass m. Furthermore, the function $R_a(k)$ does not vanish at the pole value $k^2 = m_a^2$. The above equation can then be written as

$$\tfrac{1}{2}\Delta'_{aF}(x_a,m_a) = -i(\partial_a^2 + m_a^2)^{-1}R_a\left(i\frac{\partial}{\partial x_a}\right)\delta^4(x_a).$$

Thus the factors $R_l(i(\partial/\partial x_l))$ and $R_2(i(\partial/\partial x_2))$ of particles 1 and 2, respectively, can be included, in multiplicative form, in the usual BS kernel

$$K\left(y_1,y_2,x_1',x_2';i\frac{\partial}{\partial x_1},i\frac{\partial}{x_2}\right) \equiv K_0(y_1,y_2,x_1',x_2')R_1\left(i\frac{\partial}{\partial x_1}\right)R_2\left(i\frac{\partial}{\partial x_2}\right) \equiv K_0(y_1,y_2,x_1',x_2')R\left(i\frac{\partial}{\partial X},i\frac{\partial}{\partial x}\right).$$

We notice that Eqs. (2.11) are also valid for scattering systems.

Because of translation invariance, the kernel K actually depends on three relative coordinate four-vectors and we write it in the form[28]

$$K = K\left(y_1-x_1',y_2-x_2',x_1'-x_2';i\frac{\partial}{\partial X},i\frac{\partial}{\partial x}\right) \equiv K(z_1,z_2,x';p,v), \qquad (2.13)$$

and we introduce the notation

$$z_1 = y_1 - x_1', \quad z_2 = y_2 - x_2', \quad Z = \tfrac{1}{2}(z_1+z_2), \quad z = z_1 - z_2, \quad X' = \tfrac{1}{2}(x_1'+x_2'), \quad x' = x_1' - x_2', \quad w = x - x' - z. \qquad (2.14)$$

Furthermore the wave function Φ, being an eigenfunction of p_μ, can be written as

$$\Phi(x_1,x_2) = e^{-ip\cdot X}\phi(x). \qquad (2.15)$$

We now take the difference and the sum of the two equations (2.11). After making the change of variables (2.14) and using (2.15) we get

$$p\cdot v\Phi(X,x)=\tfrac{1}{2}(m_1^2-m_2^2)\Phi(X,x) -\frac{i}{2}\int d^4Z\,d^4z\,d^4x'\,K(z_1,z_2,x';p,v) \times e^{ip\cdot Z}\left[\tfrac{1}{2}\Delta_{2F}(-w)e^{-ip\cdot w/2} -\tfrac{1}{2}\Delta_{1F}(w)e^{ip\cdot w/2}\right]\Phi(X,x'), \tag{2.16}$$

$$\left[\frac{1}{4}p^2+v^{T2}+\frac{(p\cdot v)^2}{p^2}-\frac{1}{2}(m_1^2+m_2^2)\right]\Phi(X,x) =-\frac{i}{2}\int d^4Z\,d^4z\,d^4x'\,K(z_1,z_2,x';p,v) \times e^{ip\cdot Z}\left[\tfrac{1}{2}\Delta_{2F}(-w)e^{-ip\cdot w/2} +\tfrac{1}{2}\Delta_{1F}(w)e^{ip\cdot w/2}\right]\Phi(X,x'). \tag{2.17}$$

Equation (2.16) can be used to find the function Ψ that satisfies Eq. (2.7). By using representation (2.12) and rearranging certain terms one finds

$$\Psi(X,x)\equiv\Phi(X,x)+\frac{i}{4(p^2)^{1/2}}\int d^4Z\,d^4z\,d^4x'\,K(z_1,z_2,x';p,v)e^{ip\cdot Z} \times\Bigg\{\frac{1}{b_1(b_1+a+c)}\left[\theta(w_L)e^{-i(b_1-a)w_L}+\theta(-w_L)e^{i(b_1+a)w_L}\right] +\frac{1}{b_2(b_2+a-c)}\left[\theta(w_L)e^{-i(b_2+a)w_L}+\theta(-w_L)e^{i(b_2-a)w_L}\right] +\frac{2e^{-icw_L}}{(a^2+c^2-\frac{1}{2}(b_1^2+b_2^2))}\left[\theta(w_L)(e^{-i(b_1-a-c)w_L}-1)+\theta(-w_L)(e^{i(b_2-a+c)w_L}-1)\right]\Bigg\}\delta^3(w^T)\Phi(X,x'), \tag{2.18}$$

where we have defined

$$a=(p^2)^{1/2}/2,\quad b_1=(m_1^2-v^{T2})^{1/2},\quad b_2=(m_2^2-v^{T2})^{1/2},\quad c=(m_1^2-m_2^2)/2(p^2)^{1/2}. \tag{2.19}$$

If the BS wave function Φ belongs to the sector of solutions of the BS equation which have nonrelativistic limits, then the operations defined inside the brackets in relation (2.18) are well defined and have nonrelativistic limits.

We shall write Eq. (2.18) with the symbolic notation

$$\Psi\equiv\Phi+K_1*\Phi. \tag{2.20}$$

In perturbation theory, and hence for the sector of normal solutions of the BS equation,[11] this relation can be inverted and Φ can be expressed in terms of Ψ:

$$\Phi(X,x)=(1+K_1*)^{-1}\Psi(X,x) \equiv\Psi(X,x)+\int d^4x'\,K_2(x,x';p)\Psi(X,x'). \tag{2.21}$$

Next, we replace the operator $(p\cdot v)^2$ in Eq. (2.17) by its expression obtained from Eq. (2.16), on which we apply the operator $p\cdot v$. After using representation (2.12), we bring some of the terms of the right-hand side of Eq. (2.17) to the left, to make the function Ψ (2.18) appear. We then get the equation

$$\left[\frac{1}{4}p^2-\frac{1}{2}(m_1^2+m_2^2)+\frac{(m_1^2-m_2^2)^2}{4p^2}+v^{T^2}\right]\Psi(X,x) =-\frac{i}{2(p^2)^{1/2}}\int d^4Z\,d^4z\,d^4x'\,K(z_1,z_2,x';p\cdot v) \times e^{ip\cdot Z}\delta^3(w^T)e^{-icw_L}\Phi(X,x'). \tag{2.22}$$

The right-hand side has the same x_L dependence (2.8) of the function Ψ of the left-hand side as it should be.

Replacing Φ by its expression (2.21) in terms of Ψ we obtain the second equation satisfied by Ψ:

$$\left[\frac{1}{4}p^2-\frac{1}{2}(m_1^2+m_2^2)+\frac{(m_1^2-m_2^2)^2}{4p^2}+v^{T2}\right]\Psi(X,x) =-\frac{i}{2(p^2)^{1/2}}\int d^4Z\,d^4z\,d^4x'\,K(z_1,z_2,x';p,v) \times e^{ip\cdot Z}\delta^3(w^T)e^{-icw_L}(1+K_1*)^{-1}\Psi(X,x'), \tag{2.23}$$

which is to be compared to Eq. (2.10) [the half-sum of the two equations (2.10)]. By identifying the right-hand sides of both equations [(2.10) and (2.23)] we get the expression for the potential V in terms of the kernel of the BS equation:

$$V(x^T,x'^T;p)=-\frac{i}{2(p^2)^{1/2}}\int d^4Z\,d^4z\,d^4x''dx'_L\,K(z_1,z_2,x'';p,v_L=c,v^T) \times e^{ip\cdot Z}\delta^3(x^T-x''^T-z)e^{-ic(x'_L-x''_L-z_L)}\left[\delta^4(x''-x')+K_2(x'',x';p)\right]. \tag{2.24}$$

Equations (2.18)–(2.24) determine, in the spin-0 case, the relationship between the quantum mechanical wave function and its interaction potential on the one hand, and the corresponding BS wave function and the BS kernel on the other, for the sector of solutions that have nonrelativistic limits.

We end this section by noticing that relation (2.18) can also be expressed by means of the propagator functions, after using representation (2.12) (Ref. 27); this renders the graphical interpretation of the corresponding iteration series in perturbation theory more transparent:

$$\Psi(X,x) \equiv \Phi(X,x) + \frac{i}{4a}\int d^4Z\, d^4z\, d^4x'\, K(z_1,z_2,x';p,v)$$
$$\times e^{ip\cdot Z}\left\{\frac{e^{iaw_L}}{(b_1+a+c)}\left[1 - \frac{(1-e^{i(b_1-a-c)w_L})}{(b_1-a-c)}\left(b_1 + i\frac{\partial}{\partial x_L}\right)\right]\frac{1}{2}\Delta_{1F}(w,m_1)\right.$$
$$\left. + \frac{e^{-iaw_L}}{(b_2+a-c)}\left[1 - \frac{(1-e^{-i(b_2-a+c)w_L})}{(b_2-a+c)}\left(b_2 - i\frac{\partial}{\partial x_L}\right)\right]\frac{1}{2}\Delta_{2F}(-w,m_2)\right\}\Phi(X,x'), \tag{2.25}$$

or

$$\Phi(x_1,x_2) = \Psi(x_1,x_2) - \frac{i}{4a}\int d^4y_1\, d^4y_2\, d^4x_1'\, d^4x_2'\, K(y_1-x_1',y_2-x_2',x';p,v)$$
$$\times\left\{\frac{1}{(b_1+a+c-i\epsilon)}\left[1 - \frac{(1-e^{i(b_1-a-c)(x-y)_L})}{(b_1-a-c-i\epsilon)}(b_1+i\partial_{1L})\right]\delta^4(x_2-y_2)\frac{1}{2}\Delta_{1F}(x_1-y_1,m_1)\right.$$
$$+\frac{1}{(b_2+a-c-i\epsilon)}\left[1 - \frac{(1-e^{-i(b_2-a+c)(x-y)_L})}{(b_2-a+c-i\epsilon)}(b_2+i\partial_{2L})\right]$$
$$\left.\times\delta^4(x_1-y_1)\frac{1}{2}\Delta_{2F}(x_2-y_2,m_2)\right\}\Phi(x_1',x_2'). \tag{2.26}$$

III. THE RELATIVISTIC INSTANTANEOUS APPROXIMATION

The relations between the two sets of quantities (Ψ, V) and (Φ, K) can, in principle, be exactly evaluated in perturbation theory. This implies the occurrence of two kinds of series: the first one appearing in the evaluation of the kernel K itself, the second one in the evaluation of the inverse of the operator $(1 + K_1 *)$ in (2.21).

The lowest-order approximation of the perturbation expansion would correspond to the ladder approximation of the kernal K:

$$K(z_1,z_2,x';p,v) = \delta^4(z_1)\delta^4(z_2)D(x',p_1,p_2), \tag{3.1}$$

where the eventual dependence of D on the momenta comes from the couplings of the mediating field to the external particles.[29] In this approximation one also neglects the operator K_1 in (2.20). One then gets

$$\Phi = \Psi,$$
$$V(x^T,p_1,p_2) = -\frac{i}{2(p^2)^{1/2}}\int dx_L\, D(x_L,x^T,p_1,p_2), \tag{3.2}$$

which is nothing but the covariant form of the old instantaneous approximation.[1,19]

We shall not pursue this procedure here, since, to higher orders, the expressions become rather complicated and explicitly dependent on the type of the interaction Lagrangian.

In order to have a clearer insight into the contributions of the higher-order terms in the iteration series, we shall now appeal to an approximation that considerably simplifies the structure of the equations. We call it "the relativistic instantaneous approximation," for it appears to be a generalization of the old instantaneous approximation. It consists in replacing the various kernel operators appearing in the integrals by their mean values with respect to the covariant relative time variables (more precisely the longitudinal components of the relative coordinates), concentrated at the origin of these variables. Thus the operator $K(z_1,z_2,x';p\cdot v)$ will undergo the approximation

$$K(z_1,z_2,x';p,v) \to \delta(z_{1L})\delta(z_{2L})\delta(x_L')\int dz_{1L}\, dz_{2L}\, dx_L'$$
$$\times K(z_1,z_2,x';p,v). \tag{3.3}$$

Besides the simplifications it results in, this approximation has several other advantages. First, it is manifestly covariant and hence it preserves the relativistic invariance of the theory, considered as that of a system of two particles.[30] Second, it does not necessitate any truncation of the interaction kernel and therefore can be applied to any order of the accuracy of the evaluation of the latter. Third, it permits a partial summation of the iteration series involved in the exact relations. Fourth, it automatically eliminates the eventually existing abnormal solutions of the BS equation. This phenomenon is very analogous to the approximation, in one-dimensional quantum mechanics, of potential wells by δ-functions, in which case only the ground state solution (in an approximate form) is retained. Here, of course, the sector of ground state solutions in the relative time variable just corresponds to the sector of normal solutions of the BS equation. Furthermore, as is the case with the instantaneous approximation, this approximation yields, in the nonrelativistic limit, the exact expressions of the various quantities.[31]

As we emphasized in the Introduction, this approximation is close in spirit to the method introduced by Blankenbecler and Sugar[13] in the framework of the quasipotential approach to reduce the BS equation to a three-dimensional one and to simplify its treatment.

When approximation (3.3) is made, Eqs. (2.18) and (2.22) become, respectively,

$$\Psi(X,x)=\Phi(X,x)+\frac{i}{4(p^2)^{1/2}}\int d^4Z\,d^4z\,d^4x'\,K(z_1,z_2,x';p,v)\Big\{\frac{1}{b_1(b_1+a+c)}\left[\theta(x_L)e^{-i(b_1-a)x_L}+\theta(-x_L)e^{i(b_1+a)x_L}\right]$$
$$+\frac{1}{b_2(b_2+a-c)}\left[\theta(x_L)e^{-i(b_2+a)x_L}+\theta(-x_L)e^{i(b_2-a)x_L}\right]+\frac{2e^{-icx_L}}{(a^2+c^2-\frac{1}{2}(b_1^2+b_2^2))}$$
$$\times\left[\theta(x_L)(e^{-i(b_1-a-c)x_L}-1)+\theta(-x_L)(e^{i(b_2-a+c)x_L}-1)\right]\Big\}\delta^3(x^T-x'^T-z^T)\Phi(X,x'_L=0,x'^T), \tag{3.4}$$

$$\left[\frac{1}{4}p^2-\frac{1}{2}(m_1^2+m_2^2)+\frac{(m_1^2-m_2^2)^2}{4p^2}+v^{T2}\right]\Psi(X,x)$$
$$=-\frac{i}{2(p^2)^{1/2}}\int d^4Z\,d^4z\,d^4x'\,K(z_1,z_2,x';p,v)\delta^3(x^T-x'^T-z^T)e^{-icx_L}\Phi(X,x'_L=0,x'^T). \tag{3.5}$$

We now factorize in K the multiplicative term coming from the radiative corrections of the external propagators:

$$K(z_1,z_2,x';p,v)=R(p,v)K_0(z_1,z_2,x'). \tag{3.6}$$

We also define

$$\widetilde{R}\equiv R(p,v_L=c,v^T). \tag{3.7}$$

Equations (3.4) and (3.5) become

$$\Psi(X,x)=\Phi(X,x)+\frac{i}{4(p^2)^{1/2}}R\{\cdots\}\int d^4Z\,d^4z\,d^4x'\,K_0(z_1,z_2,x')\delta^3(x^T-x'^T-z^T)\Phi(X,x'_L=0,x'^T), \tag{3.8}$$

$$\widetilde{R}^{-1}\left[\frac{1}{4}p^2-\frac{1}{2}(m_1^2+m_2^2)+\frac{(m_1^2-m_2^2)^2}{4p^2}+v^{T2}\right]\Psi(X,x)$$
$$=-\frac{i}{2(p^2)^{1/2}}\int d^4Z\,d^4z\,d^4x'\,K_0(z_1,z_2,x')\delta^3(x^T-x'^T-z^T)e^{-icx_L}\Phi(X,x'_L=0,x'^T), \tag{3.9}$$

the dots in the curly brackets of Eq. (3.8) representing the terms in the curly brackets of Eq. (3.4).

We can now replace the integral in Eq. (3.8) by its equivalent expression given by the left-hand side of Eq. (3.9). After some algebra we get

$$\Phi(X,x)=R\widetilde{R}^{-1}\Big\{\frac{1}{2b_1}(a+c+b_1)e^{-i(b_1-a-c)x_L}\theta(x_L)+\frac{1}{2b_1}(a+c-b_1)e^{i(b_1+a+c)x_L}\theta(-x_L)$$
$$+\frac{1}{2b_2}(a-c-b_2)e^{-i(b_2+a-c)x_L}\theta(x_L)+\frac{1}{2b_2}(a-c+b_2)e^{i(b_2-a+c)x_L}\theta(-x_L)\Big\}\Psi(X,x). \tag{3.10}$$

After applying the operator v_L, contained in $R(p,v_L,v^T)$, on the variable x_L appearing on its right in the above equation [also note that the x_L dependence of Ψ is given by $\exp(-icx_L)$] one can take the limit $x_L=0$ and then replace the function $\Phi(X,x_L=0,x^T)$ in terms of Ψ in (3.9) to get the dynamical equation of Ψ. For the sake of simplicity we shall, however, neglect henceforward the contributions of the radiative corrections of the R terms and present the results without these terms.

Relation (3.10) then becomes

$$\Phi(X,x)=\Big\{\frac{1}{2b_1}(a+c+b_1)e^{-i(b_1-a-c)x_L}\theta(x_L)+\frac{1}{2b_1}(a+c-b_1)e^{i(b_1+a+c)x_L}\theta(-x_L)$$
$$+\frac{1}{2b_2}(a-c-b_2)e^{-i(b_2+a-c)x_L}\theta(x_L)+\frac{1}{2b_2}(a-c+b_2)e^{i(b_2-a+c)x_L}\theta(-x_L)\Big\}\Psi(X,x), \tag{3.11}$$

or, in terms of the propagator functions [Eq. (2.12) (Ref. 27)],

$$\Phi(X,x)=\int d^3x'^T\Big\{e^{i(a+c)x_L}\left[a+c+i\frac{\partial}{\partial x_L}\right]\frac{1}{2}\Delta_{1F}(x-x'^T,m_1)$$
$$+e^{-i(a-c)x_L}\left[a-c-i\frac{\partial}{\partial x_L}\right]\frac{1}{2}\Delta_{2F}(-x+x'^T,m_2)\Big\}\Psi(X,x_L,x'^T). \tag{3.11'}$$

For the value $x_L=0$, relation (3.11) gives

$$\Phi(X,x_L=0,x^T)=\left(\frac{a+c}{2b_1}+\frac{a-c}{2b_2}\right)\Psi(X,x_L=0,x^T)$$
$$=\int d^3x'^T\Big\{(a+c)\frac{1}{2}\Delta_{1F}(x^T-x'^T,m_1)+(a-c)\,\frac{1}{2}\,\Delta_{2F}(-x^T+x'^T,m_2)\Big\}\Psi(X,x_L=0,x'^T). \tag{3.12}$$

After replacing this function in the integral in (3.9), one gets the wave equation for Ψ:

$$\left[\frac{1}{4}p^2 - \frac{1}{2}(m_1^2 + m_2^2) + \frac{(m_1^2 - m_2^2)^2}{4p^2} + v^{T2}\right]\Psi(X,x)$$
$$= -\frac{i}{2(p^2)^{1/2}}$$
$$\times \int d^4Z\, d^4z\, d^4x'\, K(z_1,z_2,x')\delta^3(x^T - x'^T - z^T)$$
$$\times\left(\frac{a+c}{2b'_1} + \frac{a-c}{2b'_2}\right)\Psi(X,x_L,x'^T), \quad (3.13)$$

where b'_1 and b'_2 are defined as in (2.19), but with v^T replaced by v'^T (i.e., acting on x'^T).

Comparison with Eqs. (2.10) yields the expression of the potential V in terms of the kernel K:

$$V(x^T,x'^T;p) = -\frac{i}{2(p^2)^{1/2}}\int d^4Z\, d^4z\, dx'_L\, K(z_1,z_2,x')$$
$$\times\delta^3(x^T - x'^T - z^T)\left(\frac{a+c}{2b'_1} + \frac{a-c}{2b'_2}\right). \quad (3.14)$$

This equation, together with Eq. (3.12), can also be represented in the form

$$\int V(x^T,x'^T;p)\Psi(X,x_L,x'^T)d^3x'^T$$
$$= -\frac{i}{2(p^2)^{1/2}}\int d^4Z\, d^4z\, d^4x'\, K(z_1,z_2,x')$$
$$\times\delta^3(x^T - x'^T - z^T)e^{-icx_L}\Phi(X,x_L = 0,x'^T). \quad (3.15)$$

It is convenient now to define the "effective" kernel $\widetilde{K}$ by the equation

$$\widetilde{K}(x^T,x'^T)$$
$$= \int d^4Z\, d^4z\, dx'_L\, K(z_1,z_2,x')\delta^3(x^T - x'^T - z^T), \quad (3.16)$$

the expression of which in the ladder approximation (3.1) is

$$\widetilde{K}(x^T,x'^T) = \delta^3(x^T - x'^T)\widetilde{D}(x^T), \quad (3.17)$$

with

$$\widetilde{D}(x^T) = \int dx_L\, D(x_L,x^T). \quad (3.18)$$

Notice that the function $\widetilde{D}$ above may also contain momentum dependences coming from the coupling of the mediating field with the external particles.

With notation (3.16), Eqs. (3.13)–(3.15) take the simpler forms

$$\left[\frac{1}{4}p^2 - \frac{1}{2}(m_1^2 + m_2^2) + \frac{(m_1^2 - m_2^2)^2}{4p^2} + v^{T2}\right]\Psi(X,x)$$
$$= -\frac{i}{4a}\int d^3x'^T\, \widetilde{K}(x^T,x'^T)\left(\frac{a+c}{2b'_1} + \frac{a-c}{2b'_2}\right)$$
$$\times\Psi(X,x_L,x'^T), \quad (3.19)$$

$$V(x^T,x'^T;p) = -\frac{i}{4a}\widetilde{K}(x^T,x'^T)\left(\frac{a+c}{2b'_1} + \frac{a-c}{2b'_2}\right), \quad (3.20)$$

$$\int V(x^T,x'^T;p)\Psi(X,x_L,x'^T)d^3x'^T$$
$$= -\frac{i}{4a}\int \widetilde{K}(x^T,x'^T)e^{-icx_L}\Phi(X,x_L = 0,x'^T)d^3x'^T. \quad (3.21)$$

Equations (3.11) and (3.14) [or (3.20)] give the relationships between the two sets of quantities (Ψ,V) and (Φ,K) in the relativistic instantaneous approximation. Compared to the general formulas (2.18) and (2.24) they appear to be much simpler, but still keeping contributions of high-order terms in the iteration series. The remarkable feature with relation (3.11) is that it no longer makes any explicit reference to the kernel K; the effects of the operator K_1 in (2.21) have been replaced by kinematic integral operators, irrespective of the form of K. Similarly Eq. (3.20) provides a more transparent relationship between V and K, where the effects of the operator K_1 have been replaced by the kinematic integral operator $((a+c)/2b'_1 + (a-c)/2b'_2)$.

A. How to get local potentials

We now search for the kinds of approximations needed in order to have for the potential V a local function (i.e., not an integral operator). It is these kinds of potentials that are most commonly used in practical calculations in quantum mechanics and very often they provide the main physical features of the problems under study. The first approximation to be considered is, of course, the ladder approximation of the kernel K [(3.1) and (3.17)]. Here the function $D(x)$ should not be necessarily regarded as the lowest-order expression of the propagator of the mediating field, but could even represent its full propagator. More generally it might also represent an effective local approximation of the kernel K involving a series of high-order diagrams.

With approximation (3.17) relation (3.20) becomes

$$V(x^T,x'^T;p_1,p_2)$$
$$= -(i/4a)\delta^3(x^T - x'^T)\widetilde{D}(x^T;p_1,p_2)$$
$$\times((a+c)/2b'_1 + (a-c)/2b'_2). \quad (3.22)$$

We see that even in the ladder approximation the quantum mechanical potential V is still a nonlocal operator in x^T, due to the presence of the integral operators $((a+c)/2b_1 + (a-c)/2b_2)$ in Eq. (3.22). In the old instantaneous approximation (3.2) this term is absent. The potential V becomes a local function in x^T only after the latter integral operator is approximated by some local function. The simplest approximation would correspond to its replacement by a constant mean value of the type

$$\frac{1}{(m_a^2 - v^{T2})^{1/2}} \simeq \frac{1}{(m_a^2 - \langle v^{T2}\rangle)^{1/2}} \equiv \frac{1}{\langle b_a\rangle} \quad (a = 1,2), \quad (3.23)$$

the mean value being taken either as the same one for all the states (therefore representing an order of magnitude) or representing the true mean value (calculated, for instance, in the L_2 norm in the c.m. frame) for each state.

With approximation (3.23) V becomes a local function in x^T:

$$V(x^T,x'^T;p_1,p_2) = -(i/4a)\delta^3(x^T - x'^T)\widetilde{D}(x^T;p_1,p_2) \times \{(a+c)/2\langle b_1\rangle + (a-c)/2\langle b_2\rangle\}. \quad (3.24)$$

It is to be emphasized that, even in this simplified version, the relativistic instantaneous approximation provides nontrivial corrections coming from the iteration series of the operator K_1 (2.20). The correction term of the last parentheses becomes mostly crucial in the case of light (nearly massless) bound states, where the numerator tends to 0, while the denominator, because of the quantity $\langle v^{T2}\rangle$, remains finite.[32]

In sum, to have local potentials in quantum mechanics, one needs the following three successive approximations of the field theoretic quantities: (i) the ladder approximation of the BS kernel, (ii) the relativistic instantaneous approximation of the kernel operator, (iii) a mean value approximation of kinematic integral operators.

B. Nonrelativistic limit

We next examine the relationship between the BS wave function and the quantum mechanical wave function in two different limiting cases: the free limit and the nonrelativistic limit.

It is evident from the general relation (2.18) that in the free case (i.e., when $K=0$), Ψ and Φ are identical,

$$\Psi^{\text{free}} = \Phi^{\text{free}}. \quad (3.25)$$

We could also check whether this relationship remains true after the relativistic instantaneous approximation has been used. In the free case Ψ satisfies the wave equation

$$\left[\frac{1}{4}p^2 - \frac{1}{2}(m_1^2+m_2^2) + \frac{(m_1^2-m_2^2)^2}{4p^2} + v^{T^2}\right]\Psi^{\text{free}} = 0$$
$$\equiv (a+c+b_1)(a+c-b_1)\Psi^{\text{free}} = 0$$
$$\equiv (a-c+b_2)(a-c-b_2)\Psi^{\text{free}} = 0. \quad (3.26)$$

Because the physical Hilbert state is determined by the range of values $\hat{p}\cdot p_1 > 0$ and $\hat{p}\cdot p_2 > 0$ (see Ref. 10, Sec. III), the factors $(a+c+b_1)$ and $(a-c+b_2)$ are positive. Using then Eqs. (3.26) in (3.11) one finds again relation (3.25).

To study the nonrelativistic limit, we use the following expansions (here, c is the velocity of light):

$$p_0 = Mc^2 + E + \mathbf{p}^2/2M + O(c^{-2}), \quad M = m_1 + m_2,$$
$$(p^2)^{1/2} = Mc^2 + E + O(c^{-2}), \quad x_L = ct + O(c^{-1}),$$
$$x_{a0} = ct_a \ (a=1,2), \quad t = t_1 - t_2, \quad x_0^T = O(c^{-2}), \quad (3.27)$$
$$x_i^T = x_i - t(p_i/M) + O(c^{-2}), \quad v_0^T = O(c^0),$$
$$v_i^T = cu_i + O(c^{-1}), \quad \mathbf{u} = (m_2\mathbf{p}_1 - m_1\mathbf{p}_2)/M.$$

We also define the nonrelativistic c.m. variables

$$\mathbf{X}_{\text{c.m.}} = \frac{m_1\mathbf{x}_1 + m_2\mathbf{x}_2}{M}, \quad T_{\text{c.m.}} = \frac{m_1t_1 + m_2t_2}{M},$$
$$\mu = m_1/m_2/M. \quad (3.28)$$

With these expansions, relation (3.11) becomes, to leading order,

$$\Phi(X,x) = \left\{\theta(t)\exp\left[-\frac{im_2}{M}\left(\frac{\mathbf{u}^2}{2\mu} - E\right)t\right] + \theta(-t)\exp\left[i\frac{m_1}{M}\left(\frac{\mathbf{u}^2}{2\mu} - E\right)t\right]\right\}\Psi(X,x). \quad (3.29)$$

[Notice that the arguments of the exponentials in (3.11) contain a factor $(\hbar c)^{-1}$; also in the nonrelativistic limit the exponentials $e^{i(b_1+a+c)x_L}$ and $e^{-i(b_2+a-c)x_L}$ disappear by rapid oscillations.] Relation (3.29) shows that, even in the nonrelativistic limit, the wave functions Φ and Ψ are not identical.

To get the full expression of the wave function Φ in terms of the internal wave function $\psi(x^T)$, we also expand the function Ψ. Using expression (2.8) and the variables (3.28) one ends up with the formula

$$\Phi(x_1,x_2) = \exp[-i(Mc^2 + \mathbf{p}^2/2M + E)T_{\text{c.m.}}] \times \exp[-i\mathbf{p}\cdot\mathbf{X}_{\text{c.m.}}]\left\{\theta(t)\exp\left[-i\left(\frac{\mathbf{u}^2}{2m_1} - \frac{m_1}{M}E\right)t\right] + \theta(-t)\exp\left[i\left(\frac{\mathbf{u}^2}{2m_2} - \frac{m_2}{M}E\right)t\right]\right\}\psi\left(\mathbf{x} - \frac{\mathbf{p}}{M}t\right). \quad (3.30)$$

For bound state problems (in nonconfining interactions) E is negative. We therefore deduce that for positive values of the relative time variable t, the internal part of Φ has positive frequencies and for negative values of t it has negative frequencies, in accordance with the general properties of the BS wave function, as first demonstrated by Wick.[4]

As far as the potential is concerned, it reduces from relation (3.22), or even from (3.20), to the usual static potential obtained with the old instantaneous approximation in the extreme nonrelativistic limit.[1,19] [Its covariant form is given in this case by (3.2).]

C. Analyticity property in the relative energy variable

To end this section we shall examine the impact of the relativistic instantaneous approximation on two theoretical properties of the BS wave function. The first one concerns its analyticity property in momentum space of the relative energy variable. The second one concerns the causality property of the commutator part of the BS wave function.

As was shown by Wick,[4] for bound state problems in nonconfining interactions, the BS wave function displays an analyticity property in the momentum space of the relative energy variable. Stated differently, the internal wave function, for positive values of the relative time variable, contains positive frequencies only, while for negative values of the relative time variable, it contains negative frequencies only. This result is obtained by inserting a complete set of states in the matrix element defining the BS wave function and by noticing that the bound states have total masses $(p^2)^{1/2}$ smaller than the sum of the masses of each constituent $[(p^2)^{1/2} < M]$.

This property was already checked previously in the nonrelativistic limit [from formula (3.30)], but we also wish to check it in the general case from formula (3.11). First, we note that the quantum mechanical wave function Ψ contains a relative time dependence of the type $\exp(-icx_L)$ (2.8), which balances the terms $\exp(icx_L)$ contained in the curly brackets in (3.11). Furthermore, using, instead of the variable X (2.5), Wick's total variable $X_{\text{c.m.}}$ (3.28), relationship (3.11) becomes

$$\Phi(x_1,x_2) = \exp[-ip\cdot X_{\text{c.m.}}]\{(1/2b_1)(a+c+b_1)e^{-i(b_1-a-cp^2/M^2)x_L}\theta(x_L) + (1/2b_1)(a+c-b_1)e^{i(b_1+a+cp^2/M^2)x_L}\theta(-x_L) + (1/2b_2)(a-c-b_2)e^{-i(b_2+a-cp^2/M^2)x_L}\theta(x_L) + (1/2b_2)(a-c+b_2)e^{i(b_2-a+cp^2/M^2)x_L}\theta(-x_L)\}\psi(x^T). \quad (3.31)$$

The argument of the first exponential in the curly brackets is (besides the factor $-ix_L$),

$$b_1 - a - c\frac{p^2}{M^2} = (m_1^2 - v^{T^2})^{1/2} - \frac{(p^2)^{1/2}}{2} - \frac{(m_1-m_2)}{2M}(p^2)^{1/2}, \quad (3.32)$$

and, since v^T is spacelike,

$$b_1 - a - c\frac{p^2}{M^2} > m_1 - \frac{m_1}{M}(p^2)^{1/2} = m_1\left(1 - \frac{(p^2)^{1/2}}{M}\right) > 0, \quad (3.33)$$

because $(p^2)^{1/2} < M$ for bound states in nonconfining interactions. Therefore the frequency for $x_L > 0$ is positive. Similar analyses can be repeated for the other terms in the curly brackets of relationship (3.31). The general result remains: the relativistic instantaneous approximation maintains the relative energy spectral properties of the BS wave function.

We now come to the causality property of the commutator part of the BS wave function. As is known, the latter is defined as the matrix element of the T product of two fields between vacuum and the state $|p\rangle$. The general spectral properties of this function are those of vertex functions satisfying the Deser–Gilbert–Sudarshan representation.[25] Instead of the T product we could consider as well the matrix element of the commutator of two fields, in which case the vertex function should display causal properties (i.e., should vanish for spacelike relative distances x). The question arises as to whether this property is also maintained by the relativistic instantaneous approximation. The answer is, in general, negative.

The phenomenon is best seen in the equal-mass case $m_1 = m_2 = m$. The relationship (3.11) can also be written in integral form

$$\Phi(X,x) = 2(2\pi)^3\left(\frac{(p^2)^{1/2}}{2}\cos\left(\frac{p\cdot x}{2}\right) - \sin\left(\frac{p\cdot x}{2}\right)\frac{\partial}{\partial x_L}\right) \times \int \Psi(X,x'^T)\,\frac{1}{2}\,\Delta_F(x-x'^T,m)\,d^3x'^T. \quad (3.34)$$

In order to obtain the expression of the "causal" wave function, given by the matrix element of the commutator of the two fields, one simply replaces the propagator function of the right-hand side of (3.34) by the Pauli-Jordan function. One gets

$$\Phi_c(X,x) = 2i(2\pi)^3\left(\frac{(p^2)^{1/2}}{2}\cos\left(\frac{p\cdot x}{2}\right) - \sin\left(\frac{p\cdot x}{2}\right)\frac{\partial}{\partial x_L}\right) \times \int \Psi(X,x'^T)\Delta(x-x'^T,m)\,d^3x'^T. \quad (3.35)$$

Except in the hyperplane $x_L = 0$ and its neighborhood, determined by the spreading radius of the wave function $\psi(x^T)$, the function $\Phi_c(X,x)$ does not in general vanish for spacelike values of x. This shows that the relativistic instantaneous approximation does not preserve, in general, the causality properties of the local fields of the underlying field theory, although the relativistic invariance of the system, considered as that of two constituent particles, is maintained [because of the manifest covariance of the approximation and the compatibility of the two wave equations (2.10), the structure of which is not altered by this approximation].

IV. SPIN-$\frac{1}{2}$ FERMION–ANTIFERMION SYSTEMS

In this section we consider systems composed of spin-$\frac{1}{2}$ one fermion and one antifermion, although their masses might be different. The treatment of two fermion systems follows similar lines.

In the manifestly covariant two-particle relativistic quantum mechanics the fermion–antifermion wave function satisfies two independent wave equations,[10] which are generalizations of the Dirac equations relative to the fermion (particle 1) and to the antifermion (particle 2) and are the analogs of Eqs. (2.1) of the spin-0 case,

$$H_1\Psi \equiv [\gamma\cdot p_1 - m_1 - (-\eta\cdot p_2 + m_2)V]\Psi = 0, \quad (4.1a)$$

$$H_2\Psi \equiv [\eta\cdot p_2 + m_2 + (\gamma\cdot p_1 + m_1)V]\Psi = 0, \quad (4.1b)$$

where the wave function Ψ is a 16 component spinor of rank 2,

$$\Psi = \Psi_{\alpha_1\alpha_2}(x_1,x_2) \quad (\alpha_1,\alpha_2 = 1,\ldots,4). \quad (4.2)$$

The matrices γ and η are the Dirac matrices acting on the fermion and antifermion spinor indices, respectively (labeled by subindices 1 and 2):

$$\gamma_\mu\Psi \equiv \gamma_{1\mu}\Psi = (\gamma_\mu)_{\alpha_1\beta_1}\Psi_{\beta_1\alpha_2}, \qquad \eta_\mu\Psi \equiv \Psi\gamma_{2\mu} = \Psi_{\alpha_1\beta_2}(\gamma_\mu)_{\beta_2\alpha_2}. \quad (4.3)$$

(The matrices γ and η commute.)

The potential V is a Poincaré invariant function of the coordinates, momenta, and Dirac matrices. The compatibility condition (2.2) of the wave equations requires that V depend on the relative coordinates x through the transverse components x^T alone[10] [(2.5)],

$$V = V(x^T, p_1, p_2, \gamma, \eta) \tag{4.4}$$

[where V satisfies Eq. (2.3)].

Equations (4.1) completely determine the longitudinal relative coordinate x_L dependence of the wave function through the equation

$$(p_1^2 - p_2^2)\Psi = (m_1^2 - m_2^2)\Psi, \tag{4.5}$$

which is a consequence of Eqs. (4.1) and the solution of which is, for eigenfunctions of the total momentum p, given by the decomposition (2.8).

As is the spin-0 case the potentials V can also be generalized so as to include nonlocal potentials in x^T (i.e., integral operators):

$$(\gamma\cdot p_1 - m_1)\Psi(X,x) = (-\eta\cdot p_2 + m_2)\int V(x^T,x'^T;\gamma,\eta;p) \times\Psi(X,x_L,x'^T)d^3x'^T, \tag{4.6a}$$

$$(\eta\cdot p_2 + m_2)\Psi(X,x) = -(\gamma\cdot p_1 + m_1)\int V(x^T,x'^T;\gamma,\eta;p) \times\Psi(X,x_L,x'^T)d^3x'^T, \tag{4.6b}$$

where V has the same properties as the one appearing in the spin-0 case (2.10) and Ψ has the same dependence on the relative longitudinal coordinate x_L as in Eq. (2.8) [it satisfies Eq. (4.5)].

In order to find the relation of these equations with the BS equation, we start again from the integrodifferential form of the latter:

$$\begin{aligned}(\gamma\cdot p_1 - m_1)\Phi(x_1,x_2) &= i(\eta\cdot p_2 - m_2)\int d^4y_1\, d^4y_2\, d^4x_1'\, d^4x_2' \\ &\times K\left(y_1,y_2,x_1',x_2';i\frac{\partial}{\partial x_1},i\frac{\partial}{\partial x_2}\right)\frac{1}{2}\,\Delta_{2F}(x_2-y_2,m_2) \\ &\times\delta^4(x_1-y_1)\Phi(x_1',x_2'),\end{aligned} \tag{4.7a}$$

$$\begin{aligned}(\eta\cdot p_2 + m_2)\Phi(x_1,x_2) &= i(\gamma\cdot p_1 + m_1)\int d^4y_1\, d^4y_2\, d^4x_1'\, d^4x_2' \\ &\times K\left(y_1,y_2,x_1',x_2';i\frac{\partial}{\partial x_1},i\frac{\partial}{\partial x_2}\right)\frac{1}{2}\,\Delta_{1F}(x_1-y_1,m_1) \\ &\times\delta^4(x_2-y_2)\Phi(x_1',x_2'),\end{aligned} \tag{4.7b}$$

where the radiative correction of the full external propagators have been included as derivative operators in the kernel K.[33]

We now multiply temporarily Eq. (4.7a) by $(\gamma\cdot p_1 + m_1)$ and Eq. (4.7b) by $(\eta\cdot p_2 - m_2)$. We get

$$\begin{aligned}(p_1^2 - m_1^2)\Phi(x_1,x_2) &= i(\gamma\cdot p_1 + m_1)(\eta\cdot p_2 - m_2)\int d(\cdots)\, K \\ &\times\tfrac{1}{2}\Delta_{2F}(x_2-y_2,m_2)\delta^4(x_1-y_1)\Phi(x_1',x_2'),\end{aligned} \tag{4.8a}$$

$$\begin{aligned}(p_2^2 - m_2^2)\Phi(x_1,x_2) &= i(\gamma\cdot p_1 + m_1)(\eta\cdot p_2 - m_2)\int d(\cdots)\, K \\ &\times\tfrac{1}{2}\Delta_{1F}(x_1-y_1,m_1)\delta^4(x_2-y_2)\Phi(x_1',x_2').\end{aligned} \tag{4.8b}$$

Compared to Eqs. (2.11), these equations have the same structure as the latter, except for the overall multiplicative factor $-(\gamma\cdot p_1 + m_1)(\eta\cdot p_2 - m_2)$. We can now repeat the same procedure as for the spin-0 case and define the wave function Ψ that must satisfy Eq. (4.5) and have the structure (2.8). The answer is

$$\Psi = \Phi - (\gamma\cdot p_1 + m_1)(\eta\cdot p_2 - m_2)K_1*\Phi, \tag{4.9}$$

where the operator K_1* is defined by Eqs. (2.18)–(2.20).

The wave function Ψ satisfies a Klein–Gordon type equation similar to Eq. (2.22), but with the right-hand side multiplied by the factor $-(\gamma\cdot p_1 + m_1)(\eta\cdot p_2 - m_2)$. To get the corresponding two Dirac-type equations, it is then sufficient to divide the equation either by $(\gamma\cdot p_1 + m_1)$ or by $(\eta\cdot p_2 - m_2)$. We finally get

$$\begin{aligned}(\gamma\cdot p_1 - m_1)\Psi(X,x) &= \frac{i}{2(p^2)^{1/2}}(\eta\cdot p_2 - m_2)\int d^4Z\, d^4z\, d^4x' \\ &\times K(z_1,z_2,x';p,v)e^{ip\cdot Z}\delta^3(w^T)e^{-icw_L}\Phi(X,x'),\end{aligned} \tag{4.10a}$$

$$\begin{aligned}(\eta\cdot p_2 + m_2)\Psi(X,x) &= \frac{i}{2(p^2)^{1/2}}(\gamma\cdot p_1 + m_1)\int d^4Z\, d^4z\, d^4x' \\ &\times K(z_1,z_2,x';p,v)e^{ip\cdot Z}\delta^3(w^T)e^{-icw_L}\Phi(X,x'),\end{aligned} \tag{4.10b}$$

with notation already introduced in (2.14) and (2.19).

In perturbation theory, and for the sector of solutions having nonrelativistic limits, Eq. (4.9) can be inverted to express Φ in terms of Ψ. Then it is replaced in the integrals in Eqs. (4.10) to get the final equations satisfied by Ψ. By identification with Eqs. (4.6) one then gets the relation between the potential V and the kernel K, in a similar way to Eq. (2.24) of the spin-0 case.

In the remaining part of this section, we shall concentrate on the relativistic instantaneous approximation (also neglecting the radiative corrections of the external propagators) in which case the above relationships become simpler and more transparent.

After approximation (3.3) is made and definition (3.16) is used,[33] Eqs. (4.9) and (4.10) become, respectively,

$$\Psi(X,x) = \Phi(X,x) - i(\gamma\cdot p_1 + m_1)(\eta\cdot p_2 - m_2)\frac{1}{4(p^2)^{1/2}}\int d^3x'^T$$
$$\times \widetilde{K}(x^T,x'^T)\Big\{\frac{1}{b_1(b_1+a+c)}\left[\theta(x_L)e^{-i(b_1-a)x_L} + \theta(-x_L)e^{i(b_1+a)x_L}\right]$$
$$+\frac{1}{b_2(b_2+a-c)}\left[\theta(x_L)e^{-i(b_2+a)x_L} + \theta(-x_L)e^{-i(b_2-a)x_L}\right]$$
$$+\frac{2e^{-icx_L}}{(a^2+c^2-\frac{1}{2}(b_1^2+b_2^2))}\left[\theta(x_L)(e^{-i(b_1-a-c)x_L}-1) + \theta(-x_L)(e^{i(b_2-a+c)x_L}-1)\right]\Big\}$$
$$\times \Phi(X,x'_L = 0,x'^T), \tag{4.11}$$

$$(\gamma\cdot p_1 - m_1)\Psi(X,x) = \frac{i}{2(p^2)^{1/2}}(\eta\cdot p_2 - m_2)\int d^3x'^T \widetilde{K}(x^T,x'^T)e^{-icx_L}\Phi(X,x'_L = 0,x'^T), \tag{4.12a}$$

$$(\eta\cdot p_2 + m_2)\Psi(X,x) = \frac{i}{2(p^2)^{1/2}}(\gamma\cdot p_1 + m_1)\int d^3x'^T \widetilde{K}(x^T,x'^T)e^{-icx_L}\Phi(X,x'_L = 0,x'^T). \tag{4.12b}$$

Comparison of Eqs. (4.12) with Eqs. (4.6) shows immediately that relation (3.21) of the spin-0 case holds also in the spin-$\frac{1}{2}$ case.

We now consider relation (4.11). Because the radiative corrections of the external propagators have been neglected, we can bring the curly brackets outside the integral. We then commute the operators $(\gamma\cdot p_1 + m_1)$ and $(\eta\cdot p_2 - m_2)$ with the expressions contained in the curly brackets and also apply the longitudinal components of p_1 and p_2 on the variable x_L. Finally we use Eqs. (4.12) and (3.21) to get the explicit relationship between Φ and Ψ:

$$\Phi(X,x) = \{(1/2b_1)(a+c+b_1)e^{-i(b_1-a-c)x_L}\theta(x_L) + (1/2b_1)(a+c-b_1)e^{i(b_1+a+c)x_L}\theta(-x_L)$$
$$+ (1/2b_2)(a-c-b_2)e^{-i(b_2+a-c)x_L}\theta(x_L) + (1/2b_2)(a-c+b_2)e^{i(b_2-a+c)x_L}\theta(-x_L)\}\Psi(X,x)$$
$$+\{(1/2b_1)(\theta(x_L)e^{-i(b_1-a-c)x_L} + \theta(-x_L)e^{i(b_1+a+c)x_L})$$
$$-(1/2b_2)(\theta(x_L)e^{-i(b_2+a-c)x_L} + \theta(-x_L)e^{i(b_2-a+c)x_L})\}$$
$$\times[\gamma\cdot\hat{p}(\eta\cdot p_2 - m_2) - \eta\cdot\hat{p}(\gamma\cdot p_1 + m_1)]\int V(x^T,x'^T;\gamma,\eta;p)\Psi(X,x_L,x'^T)d^3x'^T$$
$$+\{(1/2b_1)\left[\theta(x_L)(a+c-b_1)e^{-i(b_1-a-c)x_L} + \theta(-x_L)(a+c+b_1)e^{i(b_1+a+c)x_L}\right]$$
$$+(1/2b_2)\left[\theta(x_L)(a-c+b_2)e^{-i(b_2+a-c)x_L} + \theta(-x_L)(a-c-b_2)e^{i(b_2-a+c)x_L}\right]\}$$
$$\times\gamma\cdot\hat{p}\,\eta\cdot\hat{p}\int V(x^T,x'^T;\gamma,\eta;p)\Psi(X,x_L,x'^T)d^3x'^T. \tag{4.13}$$

In the hyperplane $x_L = 0$, this relation becomes

$$\Phi(X,x_L = 0,x^T) = \left(\frac{a+c}{2b_1} + \frac{a-c}{2b_2}\right)\Psi(X,x_L = 0,x^T) + \left(\frac{1}{2b_1} - \frac{1}{2b_2}\right)[\gamma\cdot\hat{p}(\eta\cdot\tilde{p}_2 - m_2) - \eta\cdot\hat{p}(\gamma\cdot\tilde{p}_1 + m_1)]$$
$$\times\int V(x^T,x'^T;\gamma,\eta;p)\Psi(X,x_L = 0,x'^T)d^3x'^T$$
$$+\left(\frac{a+c}{2b_1} + \frac{a-c}{2b_2}\right)\gamma\cdot\hat{p}\,\eta\cdot\hat{p}\int V(x^T,x'^T;\gamma,\eta;p)\Psi(X,x_L = 0,x'^T)d^3x'^T, \tag{4.14}$$

where we have defined

$$\tilde{p}_{1\mu} = p_{1\mu} - v_\mu^L + c\hat{p}_\mu, \quad \tilde{p}_{2\mu} = p_{2\mu} + v_\mu^L - c\hat{p}_\mu. \tag{4.15}$$

Inserting the expression (4.14) of $\Phi(X,x_L = 0,x^T)$ in relation (3.21) we get the relationship between the potential V and the kernel $\widetilde{K}$,

$$V(x^T,x'^T) = -\frac{i}{2(p^2)^{1/2}}\int d^3x''^T\widetilde{K}(x^T,x''^T)\Big\{\left(\frac{a+c}{2b''_1} + \frac{a-c}{2b''_2}\right)\delta^3(x''^T - x'^T)$$
$$+\left(\frac{1}{2b''_1} - \frac{1}{2b''_2}\right)[\gamma\cdot\hat{p}(\eta\cdot\tilde{p}''_2 - m_2) - \eta\cdot\hat{p}(\gamma\cdot\tilde{p}''_1 + m_1)]V(x''^T,x'^T)$$
$$+\left(\frac{a+c}{2b''_1} + \frac{a-c}{2b''_2}\right)\gamma\cdot\hat{p}\,\eta\cdot\hat{p}V(x''^T,x'^T)\Big\}, \tag{4.16}$$

where, for simplicity of notation, we have omitted from V and $\widetilde{K}$ the momenta and Dirac matrices; the operators b''_a $(a = 1,2)$ are defined as in relation (3.13) for b'_a.

In the ladder-type approximation (3.1) and (3.17), the previous relations simplify. Equation (4.16) becomes

$$V(x^T,x'^T) = -\frac{i}{2(p^2)^{1/2}}\widetilde{D}(x^T)\left\{\left(\frac{a+c}{2b_1}+\frac{a-c}{2b_2}\right)\delta^3(x^T-x'^T) + \left(\frac{1}{2b_1}-\frac{1}{2b_2}\right)\right.$$
$$\left.\times[\gamma\cdot\hat{p}(\eta\cdot\tilde{p}_2-m_2)-\eta\cdot\hat{p}(\gamma\cdot\tilde{p}_1+m_1)]V(x^T,x'^T)+\left(\frac{a+c}{2b_1}+\frac{a-c}{2b_2}\right)\gamma\cdot\hat{p}\,\eta\cdot\hat{p}V(x^T,x'^T)\right\}, \tag{4.17}$$

where $\widetilde{D}(x^T)$ is defined as in (3.18),

$$\widetilde{D}(x^T) = \int dx_L\, D(x_L,x^T;\gamma,\eta;p_1,p_2). \tag{4.18}$$

Also defining

$$V(x^T,x'^T) = V(x^T)\delta^3(x^T-x'^T) \tag{4.19}$$

[notice that the potential $V(x^T)$ above may still be an integral operator through operators $1/b_1$ or $1/b_2$], Eq. (4.17) takes the form

$$V = -\frac{i}{4a}\widetilde{D}(x^T)\left\{\left(\frac{a+c}{2b_1}+\frac{a-c}{2b_2}\right)(1+\gamma\cdot\hat{p}\,\eta\cdot\hat{p}V)+\left(\frac{1}{2b_1}-\frac{1}{2b_2}\right)[\gamma\cdot\hat{p}(\eta\cdot\tilde{p}_2-m_2)-\eta\cdot\hat{p}(\gamma\cdot\tilde{p}_1+m_1)]V\right\}, \tag{4.20}$$

which can be rewritten in several equivalent forms,

$$V = -\frac{i}{4a}\left[1+\frac{i}{4a}\widetilde{D}\left(\frac{1}{2b_1}-\frac{1}{2b_2}\right)(\gamma\cdot\hat{p}(\eta\cdot\tilde{p}_2-m_2)-\eta\cdot\hat{p}(\gamma\cdot\tilde{p}_1+m_1))\right.$$
$$\left.+\frac{i}{4a}\widetilde{D}\left(\frac{a+c}{2b_1}+\frac{a-c}{2b_2}\right)\gamma\cdot\hat{p}\eta\cdot\hat{p}\right]^{-1}\widetilde{D}\left(\frac{a+c}{2b_1}+\frac{a-c}{2b_2}\right), \tag{4.21a}$$

$$V = -\frac{i}{4a}\widetilde{D}\left[1+\left(\frac{1}{2b_1}-\frac{1}{2b_2}\right)(\gamma\cdot\hat{p}(\eta\cdot\tilde{p}_2-m_2)-\eta\cdot\hat{p}(\gamma\cdot\tilde{p}_1+m_1))\frac{i\widetilde{D}}{4a}\right.$$
$$\left.+\left(\frac{a+c}{2b_1}+\frac{a-c}{2b_2}\right)\gamma\cdot\hat{p}(\eta\cdot\hat{p})\frac{i\widetilde{D}}{4a}\right]^{-1}\left(\frac{a+c}{2b_1}+\frac{a-c}{2b_2}\right), \tag{4.21b}$$

$$\left[1+\left(\frac{1}{2b_1}-\frac{1}{2b_2}\right)(\gamma\cdot\hat{p}(\eta\cdot\tilde{p}_2-m_2)-\eta\cdot\hat{p}(\gamma\cdot p_1+m_1))\frac{i\widetilde{D}}{4a}+\left(\frac{a+c}{2b_1}+\frac{a-c}{2b_2}\right)\gamma\cdot\hat{p}(\eta\cdot\hat{p})\frac{i\widetilde{D}}{4a}\right]$$
$$=\left(\frac{a+c}{2b_1}+\frac{a-c}{2b_2}\right)\left[\left(\frac{a+c}{2b_1}+\frac{a-c}{2b_2}\right)(1+\gamma\cdot\hat{p}(\eta\cdot\hat{p})V)+\left(\frac{1}{2b_1}-\frac{1}{2b_2}\right)(\gamma\cdot\hat{p}(\eta\cdot\tilde{p}_2-m_2)-\eta\cdot\hat{p}(\gamma\cdot\tilde{p}_1+m_1))V\right]^{-1}. \tag{4.22}$$

Here again, the potential V, (4.21), is not a local function of x^T, because of the presence of the kinematic integral operators b_1^{-1} and b_2^{-1}, (2.19). It becomes a local function only after these integral operators are replaced by some local functions or mean values, such as in (3.23).

Also notice that, contrary to the spin-0 case, the relationship between the potential V and the kernel $\widetilde{D}$, in its ladder approximation, is not as simple as in proportionality relations, as in relations (3.22) or (3.24), unless one retains lowest-order expressions in $\widetilde{D}$ in Eq. (4.21).

Finally, with the expression (4.19) of $V(x^T, x'^T)$, the relationships (4.13) and (4.14) between the wave functions Φ and Ψ become accordingly simplified.

V. SPIN-$\frac{1}{2}$–SPIN-0 PARTICLE SYSTEMS

In this section we consider systems composed of one spin-$\frac{1}{2}$ fermion (particle 1) and one spin-0 boson (particle 2). Here the wave function is a four-component spinor

$$\Psi = \Psi_\alpha(x_1,x_2) \quad (\alpha = 1,...,4), \tag{5.1}$$

and satisfies, in the manifestly covariant two-particle relativistic quantum mechanics, two independent wave equations,[10] which are generalizations of the Dirac and Klein–Gordon equations, respectively:

$$H_1\Psi \equiv (\gamma\cdot p_1 - m_1 - V)\Psi = 0, \tag{5.2a}$$

$$H_2\Psi \equiv (p_2^2 - m_2^2 - (\gamma\cdot p_1 + m_1)V)\Psi = 0. \tag{5.2b}$$

The potential V is a Poincaré invariant function of the coordinates, momenta, and Dirac matrices. The compatibility condition (2.2) of the wave equations requires that V depend on the relative coordinates x through the transverse components x^T alone[10] [(2.5)]:

$$V = V(x^T,p_1,p_2,\gamma) \tag{5.3}$$

[V satisfies Eq. (2.3)].

Equations (5.2) completely determine the longitudinal relative coordinate x_L dependence of the wave function through the equation

$$(p_1^2 - p_2^2)\Psi = (m_1^2 - m_2^2)\Psi, \tag{5.4}$$

which is a consequence of Eqs. (5.2) and the solution of which is, for eigenfunctions of the total momentum p, given by the decomposition (2.8).

The potential V can also be generalized so as to include nonlocal potentials in x^T:

$$(\gamma\cdot p_1 - m_1)\Psi(X,x) = \int V(x^T,x'^T;\gamma;p)\Psi(X,x_L,x'^T)d^3x'^T, \tag{5.5a}$$

$$(p_2^2 - m_2^2)\Psi(X,x) = (\gamma\cdot p_1 + m_1)\int V(x^T,x'^T;\gamma;p)\Psi(X,x_L,x'^T)d^3x'^T, \tag{5.5b}$$

where V has the same properties as the one appearing in the spin-0 case (2.10) and Ψ has the same dependence on x_L as in Eq. (2.8) [it satisfies Eq. (5.4)].

In order to find the relation of these equations with the BS equation, we start again from the integrodifferential form of the latter,

$$(\gamma\cdot p_1 - m_1)\Phi(x_1,x_2)$$
$$= -i\int d^4y_1\, d^4y_2\, d^4x_1'\, d^4x_2'\, K\left(y_1,y_2,x_1',x_2';i\frac{\partial}{\partial x_1},i\frac{\partial}{\partial x_2}\right)\frac{1}{2}\Delta_{2F}(x_2-y_2,m_2)\delta^4(x_1-y_1)\Phi(x_1',x_2'), \tag{5.6a}$$

$$(p_2^2 - m_2^2)\Phi(x_1,x_2)$$
$$= -i(\gamma\cdot p_1 + m_1)\int d^4y_1\, d^4y_2\, d^4x_1'\, d^4x_2'\, K\left(y_1,y_2,x_1',x_2';i\frac{\partial}{\partial x_1},i\frac{\partial}{\partial x_2}\right)\frac{1}{2}\Delta_{1F}(x_1-y_1,m_1)\delta^4(x_2-y_2)\Phi(x_1',x_2'), \tag{5.6b}$$

with similar definitions for K as in Eqs. (2.11) and (4.7).

We now multiply temporarily Eq. (5.6a) by $(\gamma\cdot p_1 + m_1)$ and we get

$$(p_1^2 - m_1^2)\Phi(x_1,x_2) = -i(\gamma\cdot p_1 + m_1)\int d(\cdots)\,K\frac{1}{2}\Delta_{2F}(x_2-y_2,m_2)\delta^4(x_1-y_1)\Phi(x_1',x_2'), \tag{5.7a}$$

$$(p_2^2 - m_2^2)\Phi(x_1,x_2) = -i(\gamma\cdot p_1 + m_1)\int d(\cdots)\,K\frac{1}{2}\Delta_{1F}(x_1-y_1,m_1)\times\delta^4(x_2-y_2)\Phi(x_1',x_2'). \tag{5.7b}$$

Compared to Eqs. (2.11), these equations have the same structure as the latter, except for the overall multiplicative factor $(\gamma\cdot p_1 + m_1)$. We can now repeat the same procedure as for the spin-0 case and define the wave function Ψ that must satisfy Eq. (5.4) and have the structure (2.8). The answer is

$$\Psi = \Phi + (\gamma\cdot p_1 + m_1)K_1 * \Phi, \tag{5.8}$$

where the operator K_1* is defined by Eqs. (2.18)–(2.20).

The wave function Ψ satisfies a Klein–Gordon-type equation similar to Eq. (2.22), but with the right-hand side multiplied by the factor $(\gamma\cdot p_1 + m_1)$. To get the Dirac-type equation of the fermionic constituent, it is then sufficient to divide the equation by $(\gamma\cdot p_1 + m_1)$. We get

$$(\gamma\cdot p_1 - m_1)\Psi(X,x) = -\frac{i}{2(p^2)^{1/2}}\int d^4Z\, d^4z\, d^4x'\, K(z_1,z_2,x';p,v)e^{ip\cdot Z}\delta^3(w^T)e^{-icw_L}\Phi(X,x'), \tag{5.9a}$$

$$(p_2^2 - m_2^2)\Psi(X,x) = -\frac{i}{2(p^2)^{1/2}}(\gamma\cdot p_1 + m_1)\int d^4Z\, d^4z\, d^4x'\, K(z_1,z_2,x';p,v)e^{ip\cdot Z}\delta^3(w^T)e^{-icw_L}\Phi(X,x'), \tag{5.9b}$$

with notations introduced in (2.14) and (2.19).

In perturbation theory, and for the sector of solutions having nonrelativistic limits, Eq. (5.8) can be inverted to express Φ in terms of Ψ. Then it is replaced in the integrals in Eqs. (5.9) to get the final equations satisfied by Ψ. By identification with Eqs. (5.5) one then gets the relation between the potential V and the kernel K, in a similar way to Eq. (2.24) of the spin-0 case.

In the remaining part of this section, we shall concentrate on the relativistic instantaneous approximation (also neglecting the radiative corrections of the external propagators) in which case the above relationships become simpler and more transparent.

After approximation (3.3) is made and definition (3.16) is used,[33] Eqs. (5.8) and (5.9) become, respectively,

$$\Psi(X,x) = \Phi(X,x) + \frac{i}{4(p^2)^{1/2}}(\gamma\cdot p_1 + m_1)\int d^3x'^T\,\widetilde{K}(x^T,x'^T)\left\{\frac{1}{b_1(b_1+a+c)}\right.$$
$$\times\left[\theta(x_L)e^{-i(b_1-a)x_L} + \theta(-x_L)e^{i(b_1+a)x_L}\right] + \frac{1}{b_2(b_2+a-c)}\left[\theta(x_L)e^{-i(b_2+a)x_L} + \theta(-x_L)e^{i(b_2-a)x_L}\right]$$

$$+\frac{2e^{-icx_L}}{(a^2+c^2-\frac{1}{2}(b_1^2+b_2^2))}\left[\theta(x_L)(e^{-i(b_1-a-c)x_L}-1)\right.$$

$$\left.\left.+\theta(-x_L)(e^{i(b_2-a+c)x_L}-1)\right]\right\}\Phi(X,x_L'=0,x'^T), \qquad (5.10)$$

$$(\gamma\cdot p_1-m_1)\Psi(X,x)=-\frac{i}{2(p^2)^{1/2}}\int d^3x'^T\widetilde{K}(x^T,x'^T)e^{-icx_L}\Phi(X,x_L'=0,x'^T), \qquad (5.11a)$$

$$(p_2^2-m_2^2)\Psi(X,x)=-\frac{i}{2(p^2)^{1/2}}(\gamma\cdot p_1+m_1)\int d^3x'^T\widetilde{K}(x^T,x'^T)e^{-icx_L}\Phi(X,x_L'=0,x'^T). \qquad (5.11b)$$

Comparison of Eqs. (5.11) with Eqs. (5.5) shows that relation (3.21) of the spin-0 case holds also here.

We now consider relation (5.10). Because the radiative corrections of the external propagators have been neglected, we can bring the curly brackets outside the integral. We then commute the operator $(\gamma\cdot p_1+m_1)$ with the expressions contained in the curly brackets and also apply the longitudinal component of p_1 on the variable x_L. Finally we use Eqs. (5.11a) and (3.21) to get the explicit relationship between Φ and Ψ:

$$\Phi(X,x)=\{(1/2b_1)(a+c+b_1)e^{-i(b_1-a-c)x_L}\theta(x_L)+(1/2b_1)(a+c-b_1)e^{i(b_1+a+c)x_L}\theta(-x_L)$$

$$+(1/2b_2)(a-c-b_2)e^{-i(b_2+a-c)x_L}\theta(x_L)+(1/2b_2)(a-c+b_2)e^{i(b_2-a+c)x_L}\theta(-x_L)\}\Psi(X,x)$$

$$-\{(1/2b_1)\left[\theta(x_L)e^{-i(b_1-a-c)x_L}+\theta(-x_L)e^{i(b_1+a+c)x_L}\right]$$

$$-(1/2b_2)\left[\theta(x_L)e^{-i(b_2+a-c)x_L}+\theta(-x_L)e^{i(b_2-a+c)x_L}\right]\}\gamma\cdot\hat{p}\int V(x^T,x'^T;\gamma;p)\Psi(X,x_L,x'^T)d^3x'^T. \qquad (5.12)$$

In the hyperplane $x_L=0$, this relation becomes

$$\Phi(X,x_L=0,x^T)=\left(\frac{a+c}{2b_1}+\frac{a-c}{2b_2}\right)\Psi(X,x_L=0,x^T)-\left(\frac{1}{2b_1}-\frac{1}{2b_2}\right)\gamma\cdot\hat{p}\int V(x^T,x'^T;\gamma;p)\Psi(X,x_L=0,x'^T)d^3x'^T. \qquad (5.13)$$

Inserting the expression (5.13) of $\Phi(X,x_L=0,x^T)$ in relation (3.21) we get the relationship between the potential V and the kernel $\widetilde{K}$

$$V(x^T,x'^T)=-\frac{i}{2(p^2)^{1/2}}\int d^3x''^T\widetilde{K}(x^T,x''^T)$$

$$\times\left\{\left(\frac{a+c}{2b_1''}+\frac{a-c}{2b_2''}\right)\delta^3(x''^T-x'^T)\right.$$

$$\left.-\left(\frac{1}{2b_1''}-\frac{1}{2b_2''}\right)\gamma\cdot\hat{p}V(x''^T,x'^T)\right\}, \qquad (5.14)$$

where for the simplicity of notation we have omitted from V and K the momenta and Dirac matrices; the operators b_a'' $(a=1,2)$ are defined as in relation (3.13) for b_a'.

In the ladder-type approximation (3.1), (3.17), the previous relations simplify. Equation (5.14) becomes

$$V(x^T,x'^T)$$

$$=-\frac{i}{2(p^2)^{1/2}}\widetilde{D}(x^T)\left\{\left(\frac{a+c}{2b_1}+\frac{a-c}{2b_2}\right)\delta^3(x^T-x'^T)\right.$$

$$\left.-\left(\frac{1}{2b_1}-\frac{1}{2b_2}\right)\gamma\cdot\hat{p}V(x^T,x'^T)\right\}, \qquad (5.15)$$

where $\widetilde{D}(x^T)$ is defined as in relations (3.18) and (4.18). Also using definition (4.19) of (Vx^T,x'^T), Eq. (5.15) becomes

$$V=-\frac{i}{4a}\widetilde{D}(x^T)\left\{\left(\frac{a+c}{2b_1}+\frac{a-c}{2b_2}\right)\right.$$

$$\left.-\left(\frac{1}{2b_1}-\frac{1}{2b_2}\right)\gamma\cdot\hat{p}V\right\}, \qquad (5.16)$$

which gives

$$V=-\frac{i}{4a}\left[1-\frac{i}{4a}\widetilde{D}\left(\frac{1}{2b_1}-\frac{1}{2b_2}\right)\gamma\cdot\hat{p}\right]^{-1}$$

$$\times\widetilde{D}\left(\frac{a+c}{2b_1}+\frac{a-c}{2b_2}\right). \qquad (5.17)$$

The potential V becomes a local function only after the integral operators b_1^{-1} and b_2^{-1}, (2.19), are replaced by some local functions or mean values, such as in (3.23).

With the expression (4.19) of $V(x^T,x'^T)$, the relationships (5.12) and (5.13) also become simplified.

VI. THE NORMALIZATION CONDITION

For bound states, the BS equation also provides the normalization condition of the internal wave function, written as a four-dimensional integral over the relative variables.[3,20,21] This is obtained, in general, by calculating expectation values of appropriate operators (among which the charge operator) in matrix elements involving the bound state itself and thus relating them to the BS amplitude again.

If we now write the BS equation, in the spin-0 case, in its differential form, as

$$L*\Phi=(L_0-K*)\Phi=0, \qquad (6.1)$$

where

$$L_0=(p_1^2-m_1^2)(p_2^2-m_2^2) \qquad (6.2)$$

and K is the BS kernel, including the radiative corrections of the external propagators (cf. Sec. II), then the normalization condition is, in the c.m. frame,

$$-i\int d^4x\,\bar{\phi}^a(x)\frac{\partial L}{\partial p^2}*\phi(x)=1, \qquad (6.3)$$

where ϕ is the internal wave function, (2.15), and $\bar{\phi}^a$ is the "conjugate" BS wave function, obtained from ϕ by complex conjugation and antichronological product [i.e., by replacements $\theta(\pm x_0) \to \theta(\mp x_0)$].

In principle, we can express the set of quantities ϕ and K in terms of the set of quantum mechanical quantities ψ and V. Since the dependence of ψ on the longitudinal variable x_L is rather trivial, (2.8), this replacement in relation (6.3) would permit one to evaluate the x_L integration and to transform the normalization condition into a spatial three-dimensional integral, typical of quantum mechanics. This procedure, with the aid of the general formulas of the type (2.18) and (2.24), would, however, lead to rather complicated expressions, which would not be easy to handle. Instead, the use of the relativistic instantaneous approximation once again leads to compact formulas that appear to be of practical interest. It is this last procedure that we shall follow in this section, in order to express the normalization condition of the BS wave function as a three-dimensional integral for the quantum mechanical wave function ψ. We shall consider the three different systems, studied above, those involving spin-0 and/or spin-$\frac{1}{2}$ particles.

A. Two spin-0 boson systems

We begin by writing the explicit form of the operation $K*\phi$. With the definitions (2.11)–(2.15) of Sec. II, we have

$$K*\phi(x) = \int d^4Z\, d^4z\, d^4x'\, K(z_1,z_2,x';p,v)e^{ip\cdot Z} \times \delta^4(x-x'-z)\phi(x') . \quad (6.4)$$

We then calculate the expression of the operator $\partial L_0/\partial p^2$. Although formula (6.3) is valid only in the c.m. frame, we shall continue to use a covariant notation, by writing $p_\mu = \hat{p}_\mu (p^2)^{1/2}$ and deriving it by keeping $\hat{p}_\mu$ fixed. We get

$$\frac{\partial L_0}{\partial p^2} = \frac{1}{2}\left(\frac{1}{4}p^2 + v^{T^2} - \frac{1}{2}(m_1^2+m_2^2)\right) - \frac{1}{2}v_L^2 + \frac{(m_1^2-m_2^2)}{2(p^2)^{1/2}}v_L \equiv \frac{1}{2}\left(a^2 - \frac{b_1^2}{2} - \frac{b_2^2}{2}\right) - \frac{1}{2}v_L^2 + cv_L , \quad (6.5)$$

where, in the last expression, we have used the definitions (2.19).

Next, we apply this operator on the wave function $\phi(x)$, expressed in terms of $\psi(x^T)$ in the relativistic instantaneous approximation (3.11) (the radiative corrections of the external propagators have been neglected). For this, it is sufficient to evaluate the action of the operator v_L of (6.5) on the functions depending on x_L. The result is

$$\frac{\partial L_0}{\partial p^2}\phi(x) = \Big\{\tfrac{1}{2}(a+c-b_1)(a+c+b_1)\big[\theta(x_L)e^{-i(b_1-a)x_L} - \theta(-x_L)e^{i(b_1+a)x_L}\big] - \tfrac{1}{2}(a-c-b_2)(a-c+b_2)\big[\theta(x_L)e^{-i(b_2+a)x_L} - \theta(-x_L)e^{i(b_2-a)x_L}\big]\Big\}\psi(x^T) . \quad (6.6)$$

The expression of the conjugate function $\bar{\phi}^a$ is obtained from (3.11),

$$\bar{\phi}^a(x) = \psi^*(x^T)\big\{(1/2b_1)(a+c+b_1)e^{i(b_1-a)x_L}\theta(-x_L) + (1/2b_1)(a+c-b_1)e^{-i(b_1+a)x_L}\theta(x_L) + (1/2b_2)(a-c-b_2)e^{i(b_2+a)x_L}\theta(-x_L) + (1/2b_2)(a-c+b_2)e^{-i(b_2-a)x_L}\theta(x_L)\big\} , \quad (6.7)$$

where the derivative operators act on the left.

With expressions (6.6) and (6.7) we can now calculate the first term of the normalization condition (6.3). After integrating over the variable x_L and using some identities of the type

$$\frac{(b_2-a+c)}{(b_1+a+c)}(b_1+b_2+2a) = \frac{(b_1-a-c)}{(b_2+a-c)}(b_1+b_2+2a) = (b_1+b_2-2a) , \quad (6.8)$$
$$c = (b_1^2-b_2^2)/4a ,$$

we get

$$-i\int d^4x\, \bar{\phi}^a(x)\frac{\partial L_0}{\partial p^2}\phi(x) = \int d^3x^T \psi^*(x^T)\left(\frac{a+c}{2b_1} + \frac{a-c}{2b_2}\right) \times\left(a - \frac{(b_1-b_2)^2}{4a}\right)\psi(x^T) , \quad (6.9)$$

which actually is a double three-dimensional integral, because of the presence of the integral operators b_1^{-1} and b_2^{-1}.

We now turn to the calculation of the second term of the normalization condition (6.3), coming from the contribution of the BS kernel itself. Using expression (6.4) this term is

$$i\int d^4x\,\bar{\phi}^a(x)\frac{\partial K}{\partial p^2}*\phi(x) = i\int d^4Z\, d^4z\, d^4x'\, d^4x\, \bar{\phi}^a(x) \times\frac{\partial}{\partial p^2}[K(z_1,z_2,x';p,v)e^{ip\cdot Z}]\delta^4(x-x'-z)\phi(x') . \quad (6.10)$$

After using the relativistic instantaneous approximation (3.3) and neglecting the radiative corrections of the external propagators, relation (6.10) becomes

$$i\int d^4x\,\bar{\phi}^a(x)\frac{\partial K}{\partial p^2}*\phi(x) = i\int d^4Z\, d^4z\, d^4x'\, d^3x^T \bar{\phi}^a(x_L=0,x^T) \times\left(\frac{\partial}{\partial p^2}K(z_1,z_2,x')\right)\delta^3(x^T-x'^T-z^T) \times\phi(x'_L=0,x'^T)$$

$$= i \int d^3x^T d^3x'^T \overline{\phi}^a(x_L = 0,x^T) \frac{\partial \widetilde{K}}{\partial p^2} (x^T,x'^T) \times \phi(x'_L = 0,x'^T) , \quad (6.11)$$

where $\widetilde{K}$ is defined by (3.16).

Using the expression (3.12) of $\phi(x_L = 0,x^T)$ in terms of $\psi(x^T)$, we get the final expression of the normalization condition

$$(\psi,\psi) = \int d^3x^T \psi^*(x^T)\Big(\frac{a+c}{2b_1} + \frac{a-c}{2b_2}\Big) \times\Big(a - \frac{(b_1-b_2)^2}{4a}\Big)\psi(x^T) + i \int d^3x^T d^3x'^T \psi^*(x^T)\Big(\frac{a+c}{2b_1} + \frac{a-c}{2b_2}\Big) \times\frac{\partial \widetilde{K}}{\partial p^2} (x^T,x'^T)\Big(\frac{a+c}{2b'_1} + \frac{a-c}{2b'_2}\Big)\psi(x'^T) = 1 . \quad (6.12)$$

Notice that the reality condition of the norm requires from the kernel $\widetilde{K}$ to satisfy the condition

$$\widetilde{K}^+(x^T,x'^T) = -\widetilde{K}(x'^T,x^T) \quad (6.13)$$

in the L_2 norm (in the c.m. frame).

By using the relation (3.20) between V and $\widetilde{K}$ we could also make the former appear in the normalization condition. We shall, however, use this procedure in the ladder approximation where it becomes more advantageous.

In the ladder approximation (3.1), (3.17), and (3.18), the normalization condition becomes

$$(\psi,\psi) = \int d^3x^T \psi^*(x^T)\Big(\frac{a+c}{2b_1} + \frac{a-c}{2b_2}\Big) \times\Big(a - \frac{(b_1-b_2)^2}{4a}\Big)\psi(x^T) + i \int d^3x^T \psi^*(x^T)\Big(\frac{a+c}{2b_1} + \frac{a-c}{2b_2}\Big)\frac{\partial \widetilde{D}}{\partial p^2}(x^T) \times\Big(\frac{a+c}{2b_1} + \frac{a-c}{2b_2}\Big)\psi(x^T) = 1 . \quad (6.14)$$

The reality condition of the norm now requires

$$\widetilde{D}^+(x^T) = -\widetilde{D}(x^T) \quad (6.15)$$

in the L_2 norm (in the c.m. frame).

In the ladder approximation $V(x^T,x'^T)$ can be expressed as in (4.19). Using then the relationship between V and $\widetilde{D}$, (3.22), one can express $\partial\widetilde{D}/\partial p^2$ in terms of $\partial V/\partial p^2$ and V. The latter term can be eliminated by the use of the wave equations satisfied by ψ. One then gets the final result in the ladder approximation (in the c.m. frame)

$$(\psi,\psi) = \int d^3x^T \psi^*(x^T)\Big(\frac{a^2-c^2}{a}\Big)\Big(\frac{a+c}{2b_1} + \frac{a-c}{2b_2}\Big)\psi(x^T) - \int d^3x^T \psi^*(x^T) 2a\Big[\Big(\frac{a+c}{2b_1} + \frac{a-c}{2b_2}\Big)\frac{\partial V}{\partial p^2} + \frac{\partial V}{\partial p^2}\Big(\frac{a+c}{2b_1} + \frac{a-c}{2b_2}\Big)\Big]\psi(x^T) = 1 . \quad (6.16)$$

It is to be remembered that at the present stage of approximation, V is still an integral operator in x^T [cf. (3.22)]. If, furthermore, one makes the local approximation (3.23), which amounts to transforming V into a local function in x^T, then the normalization condition becomes

$$(\psi,\psi) = \Big(\frac{a+c}{\langle 2b_1\rangle} + \frac{a-c}{\langle 2b_2\rangle}\Big) \int d^3x^T \psi^*(x^T) \times\Big[\frac{a^2-c^2}{a} - 4a\frac{\partial V}{\partial p^2}(x^T,p^2)\Big]\psi(x^T) = 1 , \quad (6.17)$$

which is also equivalent to

$$\int d^3x^T \psi^*(x^T)\Big[4\hat{p}\cdot p_1(\hat{p}\cdot p_2) - 4p^2\frac{\partial V}{\partial p^2}(x^T,p^2)\Big]\psi(x^T) = 2(p^2)^{1/2}\Big(\frac{\hat{p}\cdot p_1}{2(m_1^2 - \langle v^{T2}\rangle)^{1/2}} + \frac{\hat{p}\cdot p_2}{2(m_2^2 - \langle v^{T2}\rangle)^{1/2}}\Big)^{-1} . \quad (6.18)$$

Notice that in this approximation, because of condition (6.15) and relation (3.22), V must be Hermitian in the L_2 norm (in the c.m. frame).

The interest of the last formulas (6.17) and (6.18) lies in the fact that no explicit reference is now maintained as to the field theory kernel D. If one works from the start with relativistic wave equations of the type (2.1) with effective local potentials V, then the normalization condition (6.17) and (6.18) still keeps track of the underlying (effective) field theory.

The presence of the term $\partial V/\partial p^2$ in the kernel of the norm is typical for potentials that depend explicitly on p^2 (in the c.m. frame) and this aspect was already emphasized in Ref. 34. If V does not depend on p^2 (in the c.m. frame) then the kernel of the norm (6.18) becomes simply proportional to that of the free norm of two spinless particle wave functions. The overall multiplicative factor $((a+c)/\langle 2b_1\rangle + (a-c)/\langle 2b_2\rangle)$ in the norm (6.17) is reminiscent of the field theoretic nature of the bound state $|p\rangle$. It represents a corrective factor relative to the two-constituent state inside the state $|p\rangle$, which actually is made, in interacting theories, of a series of multiconstituent states.

Although the normalization condition (6.17) has been obtained from the Bethe–Salpeter equation, which, in turn, is established only in perturbation theory and for nonconfining interactions, the fact that the validity of the relativistic wave equations (2.1) does not make any reference to the nature of the potential V (i.e., confining or nonconfining) strongly suggests that the normalization condition (6.17) might also be used for the case of confining interactions. Examples of applications for the calculation of meson decay coupling constants in quarkonium dynamics were presented in Ref. 35. It is to be emphasized that the overall factor $((a+c)/\langle 2b_1\rangle + (a-c)/\langle 2b_2\rangle)$, which will also be met, in the spin-$\frac{1}{2}$ case, plays an important role for the case of light bound states in confining interactions, the masses of which are of the order of the constituent particle masses m_1,m_2. In this case the numerator is of the order of m_1 or m_2, but the denominator is of the order of $|\langle v^{T^2}\rangle|^{1/2}$, which is governed instead by the confining interaction scale. It is the presence of this overall factor that ensured in Ref. 35 that the pion decay coupling constant not vanish in the chiral limit and hence reproduce the qualitative feature of a spontaneous breakdown of chiral symmetry. We think that this result is a positive indication to also continuing the normalization condition (6.17), or its general forms (6.12)–(6.16), to the domains of confining interactions.

A last remark concerns the positivity condition of the norm. It is well known that in relativistic quantum mechanics the shape and the strength of the interaction potentials cannot be arbitrarily chosen unless one runs into trouble with the positivity of the total mass squared p^2. We assume that the choice of V satisfies this condition. Furthermore the physical Hilbert space is chosen as corresponding to the subspace of solutions with positive eigenvalues of both $\hat{p}\cdot p_1$ and $\hat{p}\cdot p_2$.[10] We observe that this condition ensures the positivity of the norm (6.18) for p^2 independent potentials. When the potential V is p^2 dependent (in the c.m. frame), one must also control its form in order to maintain the positivity of the norm.

B. Spin-$\frac{1}{2}$ fermion–antifermion systems

Here the differential form of the BS equation is

$$L*\Phi=(L_0+K*)\Phi=0\,, \tag{6.19}$$

with

$$L_0=(\gamma\cdot p_1-m_1)(\eta\cdot p_2+m_2)\,, \tag{6.20}$$

and the normalization condition is given, in the c.m. frame, by (6.13), where

$$\overline{\phi}^a(x)=[\phi^+(x)\gamma_0\eta_0]^a\,, \tag{6.21}$$

the index a representing the antichronological product, obtained in the expression of ϕ by the exchanges $\theta(\pm x_L)\to\theta(\pm x_L)$.

The expression of the operation $K*\phi$ is still given by Eq. (6.4), with K also having dependences on the Dirac matrices. The expression of $\partial L_0/\partial p^2$ is

$$\frac{\partial L_0}{\partial p^2}=\frac{1}{4(p^2)^{1/2}}\,[\gamma\cdot\hat{p}(\eta\cdot\tilde{p}_2+m_2)+\eta\cdot\hat{p}(\gamma\cdot\tilde{p}_1-m_1)]\,, \tag{6.22}$$

where the vectors $\tilde{p}_1$ and $\tilde{p}_2$ have been defined in (4.15) (they no longer contain the longitudinal derivative operator v_μ^L). By applying the operator (6.22) on $\phi(x)$, given in the relativistic instantaneous approximation by formula (4.13), we get

$$\begin{aligned}\frac{\partial L_0}{\partial p^2}\phi(x)=&\Big\{\frac{1}{4b_1}\gamma\cdot\hat{p}(\eta\cdot\tilde{p}_2+m_2)\big[\theta(x_L)e^{-i(b_1-a)x_L}+\theta(-x_L)e^{i(b_1+a)x_L}\big]\\&+\frac{1}{4b_2}\eta\cdot\hat{p}(\gamma\cdot\tilde{p}_1-m_1)\big[\theta(x_L)e^{-i(b_2+a)x_L}+\theta(-x_L)e^{i(b_2-a)x_L}\big]\\&-\frac{(\gamma\cdot\tilde{p}_1-m_1)(\eta\cdot\tilde{p}_2+m_2)}{(a^2+c^2-\frac{1}{2}(b_1^2+b_2^2))}\Big[\frac{1}{4b_1}\big[(a+c-b_1)e^{-i(b_1-a)x_L}\theta(x_L)+(a+c+b_1)e^{i(b_1+a)x_L}\theta(-x_L)\big]\\&+\frac{1}{4b_2}\big[(a-c+b_2)e^{-i(b_2+a)x_L}\theta(x_L)+(a-c-b_2)e^{i(b_2-a)x_L}\theta(-x_L)\big]\Big]\Big\}\psi(x^T)\,.\end{aligned} \tag{6.23}$$

We have preferred to eliminate temporarily from the above expression the potential $V(x^T,x'^T)$ by means of the equations of motion (4.6). This will be more convenient for the later operation of integration in (6.3) with respect to x_L. Notice that since the dependence of the quantum mechanical wave function ψ on the variable x_L is given by formula (2.8), we can immediately replace in Eqs. (4.6) the operators p_1 and p_2 by $\tilde{p}_1$ and $\tilde{p}_2$, (4.15), respectively.

The expression of $\overline{\phi}^a$ can be obtained from (4.13). It is

$$\begin{aligned}\overline{\phi}^a(x)=&\,\overline{\psi}(x^T)\Big\{\frac{1}{2b_1}\big[(a+c+b_1)e^{i(b_1-a)x_L}\theta(-x_L)+(a+c-b_1)e^{-i(b_1+a)x_L}\theta(x_L)\big]\\&+\frac{1}{2b_2}\big[(a-c-b_2)e^{i(b_2+a)x_L}\theta(-x_L)+(a-c+b_2)e^{-i(b_2-a)x_L}\theta(x_L)\big]\Big\}\\&-\overline{\psi}(x^T)[(\gamma\cdot\tilde{p}_1-m_1)\gamma\cdot\hat{p}-(\eta\cdot\tilde{p}_2+m_2)\eta\cdot\hat{p}]\Big\{\frac{1}{2b_1}\big[e^{i(b_1-a)x_L}\theta(-x_L)+e^{-i(b_1+a)x_L}\theta(x_L)\big]\\&-\frac{1}{2b_2}\big[e^{i(b_2+a)x_L}\theta(-x_L)+e^{-i(b_2-a)x_L}\theta(x_L)\big]\Big\}-\overline{\psi}(x^T)\frac{(\gamma\cdot\tilde{p}_1-m_1)(\eta\cdot\tilde{p}_2+m_2)}{(a^2+c^2-\frac{1}{2}(b_1^2+b_2^2))}\gamma\cdot\hat{p}(\eta\cdot\hat{p})\\&\times\Big\{\frac{1}{2b_1}\big[(a+c-b_1)e^{i(b_1-a)x_L}\theta(-x_L)+(a+c+b_1)e^{-i(b_1+a)x_L}\theta(x_L)\big]\\&+\frac{1}{2b_2}\big[(a-c+b_2)e^{i(b_2+a)x_L}\theta(-x_L)+(a-c-b_2)e^{-i(b_2-a)x_L}\theta(x_L)\big]\Big\}\,,\end{aligned} \tag{6.24}$$

where the derivative operators act on the left, and again we have temporarily eliminated the potential $V(x^T,x'^T)$ by means of the equations of motion (4.6). Here $\overline{\psi}$ is defined as

$$\bar{\psi}(x^T) = \psi^+(x^T)\gamma_0\eta_0 = [\gamma_0\eta_0\psi(x^T)]^+ . \tag{6.25}$$

With the aid of formulas (6.23) and (6.24), we can now calculate the first term of the normalization condition (6.3). After integrating over the variable x_L and using identities on the type (6.8), we can transform the four-dimensional integral into a three-dimensional one over the variables x^T. The second term of the normalization condition (6.3) comes from the contribution of the term $\bar{\phi}^a(\partial K/\partial p^2)*\phi$. In the relativistic instantaneous approximation, the integral of this term has the same expression (up to a minus sign) as in relation (6.11), where the function $\phi(x_L = 0, x^T)$ now has to be replaced by its expression (4.14) in terms of $\psi(x^T)$. One finally gets

$$\begin{aligned}
(\psi,\psi) &= -i\int d^4x \,\mathrm{Tr}\left\{\bar{\phi}^a(x)\left(\frac{\partial L_0}{\partial p^2} + \frac{\partial K}{\partial p^2}*\right)\phi(x)\right\} \\
&= \int d^3x^T \,\mathrm{Tr}\,\bar{\psi}(x^T)\,\frac{1}{4a}\left\{\left(\frac{a+c}{2b_1} + \frac{a-c}{2b_2}\right)\gamma\cdot\hat{p}(\eta\cdot\hat{p})\psi(x^T)\right. \\
&\quad + \frac{1}{4a}\int d^3x^T d^3x'^T \,\mathrm{Tr}\,\bar{\psi}(x^T)\left(\frac{1}{2b_1} - \frac{1}{2b_2}\right)[\eta\cdot\hat{p}(\eta\cdot\tilde{p}_2 - m_2) - \gamma\cdot\hat{p}(\gamma\cdot\tilde{p}_1 + m_1)]V(x^T,x'^T)\psi(x'^T) \\
&\quad + \frac{1}{4a}\int d^3x^T d^3x'^T \,\mathrm{Tr}\,\bar{\psi}(x^T)\bar{V}(x^T,x'^T)[(\eta\cdot\tilde{p}_2' - m_2)\eta\cdot\hat{p} - (\gamma\cdot\tilde{p}_1' + m_1)\gamma\cdot\hat{p}]\left(\frac{1}{2b_1'} - \frac{1}{2b_2'}\right)\psi(x'^T) \\
&\quad + \frac{2c}{4a}\int d^3x^T d^3x'^T \,\mathrm{Tr}\,\bar{\psi}(x^T)\left[\left(\frac{1}{2b_1} - \frac{1}{2b_2}\right)V(x^T,x'^T) + \bar{V}(x^T,x'^T)\left(\frac{1}{2b_1'} - \frac{1}{2b_2'}\right)\right]\psi(x'^T) \\
&\quad - \frac{1}{4a}\int d^3x^T d^3x'^T d^3x''^T \,\mathrm{Tr}\,\bar{\psi}(x'^T)\bar{V}(x'^T,x^T)\gamma\cdot\hat{p}(\eta\cdot\hat{p})\left(\frac{a+c}{2b_1} + \frac{a-c}{2b_2}\right)V(x^T,x''^T)\psi(x''^T) \\
&\quad - i\int d^3x^T d^3x'^T \,\mathrm{Tr}\,\bar{\phi}^a(x_L = 0,x^T)\,\frac{\partial\tilde{K}}{\partial p^2}(x^T,x'^T)\phi(x_L' = 0,x'^T) = 1 ,
\end{aligned} \tag{6.26}$$

the trace bearing on the spinor indices, and $\bar{V}(x^T,x'^T) = \gamma_0\eta_0 V^+(x^T,x'^T)\gamma_0\eta_0$ and $\phi(x_L = 0,x^T)$, $\bar{\phi}(x_L = 0,x^T)$ are to be replaced in terms of ψ and $\bar{\psi}$ from Eq. (4.14). The relation between $\tilde{K}$ and V is given by Eq. (4.16) and the reality condition of the norm requires from $\tilde{K}$ to satisfy the relation

$$\gamma_0\eta_0\tilde{K}^+(x^T,x'^T)\gamma_0\eta_0 = -\tilde{K}(x'^T,x^T) . \tag{6.27}$$

In the ladder approximation (3.1), (3.17), and (3.18), the operator $(\partial\tilde{K}/\partial p^2)(x^T,x'^T)$ is replaced by $(\partial\tilde{D}/\partial p^2)(x^T)\delta^3(x^T - x'^T)$, and $\tilde{D}$ satisfies an analogous relation to (6.27):

$$\gamma_0\eta_0\tilde{D}^+(x^T)\gamma_0\eta_0 = -\tilde{D}(x^T) . \tag{6.28}$$

Here the relationships between $\tilde{D}$ and V are simpler [see relations (4.17)–(4.22)] and the operator $\partial\tilde{D}/\partial p^2$ may be expressed more easily in terms of V, as was done in the spin-0 case (Sec. VI A). We shall, however, present the result in the simplest case, which corresponds to the local approximation of V, where the integral operators b_1^{-1} and b_2^{-1} are replaced by mean values (3.23). After some algebra and the use of the equations of motion, one gets the final result (in the c.m. frame)

$$\begin{aligned}
(\psi,\psi) &= \frac{1}{4a}\left(\frac{a+c}{\langle 2b_1\rangle} + \frac{a-c}{\langle 2b_2\rangle}\right) \\
&\quad\times\int d^3x^T \,\mathrm{Tr}\left\{\bar{\psi}(x^T)\gamma\cdot\hat{p}(\eta\cdot\hat{p})\psi(x^T)\right. \\
&\quad - \bar{\psi}(x^T)\bar{V}(x^T,p^2)\gamma\cdot\hat{p}(\eta\cdot\hat{p})V(x^T,p^2)\psi(x^T) \\
&\quad \left. + 4p^2\bar{\psi}(x^T)\frac{\partial V}{\partial p^2}(x^T,p^2)\psi(x^T)\right\} = 1 .
\end{aligned} \tag{6.29}$$

In this approximation, V must satisfy the Hermiticity relation

$$\bar{V} \equiv \gamma_0\eta_0 V^+(x^T)\gamma_0\eta_0 = V(x^T) . \tag{6.30}$$

The presence of the potential V in the kernel of the norm, even if V is independent of p^2 in the c.m. frame, makes it clear that V must satisfy some inequality conditions to guarantee the positivity of the norm. This question was examined in more detail in Ref. 10, Sec. VII A.

The remarks made at the end of Sec. VI A for the spin-0 case in the local approximation hold also here for the spin-$\frac{1}{2}$ case. One of the advantages of formula (6.29) is its expliciit independence from field theoretical quantities, although it is derived from a relation with an underlying field theory. Notice also the appearance once again, in (6.29), of the overall factor $((a+c)/\langle 2b_1\rangle + (a-c)/\langle 2b_2\rangle)$.

C. Spin-$\frac{1}{2}$–spin-0 particle systems

The differential form of the BS equation is

$$L*\Phi = (L_0 - K*)\Phi = 0 , \tag{6.31}$$

with

$$L_0 = (\gamma\cdot p_1 - m_1)(p_2^2 - m_2^2) , \tag{6.32}$$

and the normalization condition is given, in the c.m. frame, by (6.3), where

$$\bar{\phi}^a(x) = [\phi^+(x)\gamma_0]^a , \tag{6.33}$$

the index a representing the antichronological product, obtained in the expression of ϕ by the replacements $\theta(\pm x_L) \to \theta(\mp x_L)$.

The expression of the operation $K*\phi$ is still given by Eq. (6.4), with K also having dependences on the Dirac matrices. The expression of $\partial L_0/\partial p^2$ is

$$\frac{\partial L_0}{\partial p^2} = \frac{\gamma\cdot\hat{p}}{4(p^2)^{1/2}}\Big(\frac{1}{4}p^2 + v^{T^2} - m_2^2\Big) + \frac{1}{4}\Big(\frac{1}{2}\gamma\cdot p + \gamma\cdot v^T - m_1\Big) - \frac{\gamma\cdot\hat{p}}{4(p^2)^{1/2}}v_L^2$$
$$- \Big(\frac{1}{4}\gamma\cdot\hat{p} + \frac{1}{2(p^2)^{1/2}}\gamma\cdot v^T - \frac{m_1}{2(p^2)^{1/2}}\Big)v_L\,. \tag{6.34}$$

By applying this operator on the wave function $\phi(x)$, given in the relativistic instantaneous approximation, by formula (5.12), we get

$$\frac{\partial L_0}{\partial p^2}\phi(x) = \{ -\theta(x_L)e^{-i(b_1-a)x_L}(1/4b_1)[(a+c-b_1)(\gamma\cdot\tilde{p}_1 - m_1) - \gamma\cdot\hat{p}(\tilde{p}_2^2 - m_2^2)]$$
$$-\theta(-x_L)e^{i(b_1+a)x_L}(1/4b_1)[(a+c+b_1)(\gamma\cdot\tilde{p}_1 - m_1) - \gamma\cdot\hat{p}(\tilde{p}_2^2 - m_2^2)]$$
$$-\theta(x_L)e^{-i(b_2+a)x_L}\tfrac{1}{2}(\gamma\cdot\tilde{p}_1 - m_1) + \theta(-x_L)e^{i(b_2-a)x_L}\tfrac{1}{2}(\gamma\cdot\tilde{p}_1 - m_1)\}\psi(x^T)\,, \tag{6.35}$$

where $\tilde{p}_1$ and $\tilde{p}_2$ are defined in (4.15) (they no longer contain the longitudinal derivative operator v_μ^L) and we have temporarily eliminated the potential $V(x^T,x'^T)$ through the equations of motion (5.5), as we did in Sec. VI B.

The expression of $\bar{\phi}^a$ is obtained from (5.12):

$$\bar{\phi}^a(x) = \bar{\psi}(x^T)\{(1/2b_1)[(a+c+b_1)e^{i(b_1-a)x_L}\theta(-x_L) + (a+c-b_1)e^{-i(b_1+a)x_L}\theta(x_L)]$$
$$+ (1/2b_2)[(a-c-b_2)e^{i(b_2+a)x_L}\theta(-x_L) + (a-c+b_2)e^{-i(b_2-a)x_L}\theta(x_L)]\}$$
$$-\bar{\psi}(x^T)\{(1/2b_1)[e^{i(b_1-a)x_L}\theta(-x_L) + e^{-i(b_1+a)x_L}\theta(x_L)]$$
$$-(1/2b_2)[e^{i(b_2+a)x_L}\theta(-x_L) + e^{-i(b_2-a)x_L}\theta(x_L)]\}(\gamma\cdot\tilde{p}_1 - m_1)\gamma\cdot\hat{p}\,, \tag{6.36}$$

where the derivative operators act on the left, and again we have temporarily eliminated the potential $V(x^T,x'^T)$ through the equations of motion (5.5). $\bar{\psi}$ is defined as

$$\bar{\psi}(x^T) = \psi^+(x^T)\gamma_0\,. \tag{6.37}$$

The calculation of the norm (6.3) follows now similar lines as in Secs. VI A and VI B. The result is, in the relativistic instantaneous approximation and in the c.m. frame,

$$(\psi,\psi) = -i\int d^4x\,\bar{\phi}^a(x)\Big(\frac{\partial L_0}{\partial p^2} - \frac{\partial K}{\partial p^2}*\Big)\phi(x)$$
$$= \int d^3x^T\,\bar{\psi}(x^T)\gamma\cdot\hat{p}\,\frac{1}{8ab_1}[4a^2 - (b_1-b_2)^2]\psi(x^T) - \frac{1}{2}\int d^3x^T d^3x'^T\,\bar{\psi}(x^T)\Big[\Big(\frac{1}{2b_1} - \frac{1}{2b_2}\Big)V(x^T,x'^T)$$
$$+ \bar{V}(x^T,x'^T)\Big(\frac{1}{2b_1'} - \frac{1}{2b_2'}\Big)\Big]\psi(x'^T) + i\int d^3x^T d^3x'^T\,\bar{\phi}^a(x_L=0,x^T)\,\frac{\partial\widetilde{K}}{\partial p^2}(x^T,x'^T)\phi(x_L'=0,x'^T) = 1\,, \tag{6.38}$$

where $\phi(x_L=0,x^T)$ and $\bar{\phi}(x_L=0,x^T)$ are to be replaced in terms of ψ and $\bar{\psi}$ from their expressions (5.13) and $\bar{V}(x^T,x'^T) = \gamma_0 V^+(x^T,x'^T)\gamma_0$. The relation between $\widetilde{K}$ and V is given by Eq. (5.14) and the reality condition of the norm requires from $\widetilde{K}$ to satisfy the relation

$$\gamma_0\widetilde{K}^+(x^T,x'^T)\gamma_0 = -\widetilde{K}(x'^T,x^T)\,. \tag{6.39}$$

In the ladder approximation (3.1), (3.17), and (3.18), the operator $(\partial\widetilde{K}/\partial p^2)(x^T,x'^T)$ is replaced by $(\partial\widetilde{D}/\partial p^2)(x^T)\delta^3(x^T - x'^T)$, and the operator $\widetilde{D}$ satisfies an analogous relation to (6.39):

$$\gamma_0\widetilde{D}^+(x^T)\gamma_0 = -\widetilde{D}(x^T)\,. \tag{6.40}$$

The relationships between $\widetilde{D}$ and V are given by Eqs. (5.15)–(5.17) and the operator $\partial\widetilde{D}/\partial p^2$ can be eliminated in terms of $\partial V/\partial p^2$ and V. We shall, however, present the result in the simplest case, which corresponds to the local approximation of V, where the integral operators b_1^{-1} and b_2^{-1} are replaced by mean values (3.23). After some algebra and the use of the equations of motion, the expression of the norm becomes (in the c.m. frame)

$$(\psi,\psi) = \frac{1}{4a}\Big(\frac{a+c}{2\langle b_1\rangle} + \frac{a-c}{2\langle b_2\rangle}\Big)\int d^3x^T\,\bar{\psi}(x^T)\Big[\gamma\cdot\hat{p}_2(\hat{p}\cdot p_2)$$
$$-4p^2\frac{\partial V}{\partial p^2}(x^T,p^2)\Big]\psi(x^T) = 1\,. \tag{6.41}$$

In this approximation, V must satisfy the Hermicity condition

$$\bar{V} \equiv \gamma_0 V^+(x^T)\gamma_0 = V(x^T)\,. \tag{6.42}$$

The final remark made at the end of Sec. VI A about the positivity of the norm also applies here.

VII. CONCLUSION

The main achievement of this paper is the derivation of the connection of the wave equations of two-particle relativistic quantum mechanics, based on the manifestly covariant

Hamiltonian formalism with constraints, with the Bethe–Salpeter equation. This connection establishes the link between the quantum mechanical wave function and interaction potential on the one hand, and the BS wave function and BS kernel on the other. In the quantum mechanical framework, the time components of coordinate four-vectors are treated as parameters and do not play any dynamical role, and therefore the dynamics of the system is essentially three dimensional, besides the spin degrees of freedom.

As far as the sector of normal solutions of the BS equation is concerned, i.e., the solutions having nonrelativistic limits, the above formal connection can be evaluated, in principle, in perturbation theory.

The question arises here of whether the formal connection found between the BS equation and Hamiltonian quantum mechanics also holds for the sectors of abnormal solutions of the BS equation, i.e., those not having nonrelativistic limits. We did not study this question in this paper, but we would like to mention the following two alternatives: (i) the connection does hold also for the abnormal solutions (in this case each abnormal sector would receive a corresponding Hamiltonian description, but with a specific interaction potential, not having a usual Galilei limit[24]); (ii) the connection fails (this would simply mean that the formal relations obtained so far are divergent for the abnormal sectors). Clearly this question merits a detailed investigation in the future, with the main goal of understanding whether or not the abnormal solutions may have physical existence.[3,36]

In order to simplify the relations between the quantum mechanical and field theoretical quantities, we devised, throughout this paper, an approximation, which we called the relativistic instantaneous approximation. It is a generalization of the old instantaneous approximation and it automatically selects the sector of normal solutions of the BS equation. Besides the simplifications it brings, it has the main advantage of allowing an approximate summation of the iteration series appearing in the exact relations between the quantum mechanical and field theoretical quantities. It could thus serve as a zeroth-order approximation for an expansion of the exact relations around the solution it provides. This approximation is also close in spirit to the approximations utilized in the framework of the quasipotential approach to reduce the BS equation to a three-dimensional equation and to simplify its kernel in a kind of a zeroth-order approximation of an iteration series.[13–15,18]

We showed that, in order to have, in two-particle relativistic quantum mechanics, local functions for the interaction potentials (i.e., not integral operators) three kinds of successive approximations are needed: (i) the ladder approximation of the BS kernel, (ii) the relativistic instantaneous approximation, and (iii) local or mean value approximations of some kinematic integral operators.

The connection established between the BS equation and relativistic quantum mechanics also permits one to obtain (mostly in the relativistic instantaneous approximation) the normalization condition, as a three-dimensional integral, of the quantum mechanical internal wave function. This result allows the calculation, in the quantum mechanical framework, of physical quantities such as decay coupling constants, related to the bound states under consideration.

Finally, the question we would like to raise concerns the domain of validity of the connection established so far between the BS equation and two-particle quantum mechanics. Strictly speaking, it covers perturbation theory and nonconfining interactions (this is the domain of validity of the BS equation itself). However, the structure of the wave equations of two-particle relativistic quantum mechanics is valid irrespective of the kinds of interaction potential V appearing in them, that is, whether they are confining or not. This feature strongly suggests that the relations established between field theoretic and quantum mechanical quantities, and mostly in the relativistic instantaneous approximation, where iteration series have been partly summed up, could then be continued to the domains of nonperturbative and confining interactions. In particular, the normalization condition of the quantum mechanical wave function might also be used in such cases.[35] Furthermore, the form and the structure of the quantum mechanical effective potential V used in nonperturbative interactions might also then provide some information about the structure and properties of the underlying effective field theory.

Quantum mechanics provides a suitable framework for the study of the dynamics of those problems where a finite number of degrees of freedom seems to be a good approximation. The connection established now with the Bethe–Salpeter equation brings, in this respect, additional qualitative and quantitative informations for the evaluation of the related physical quantities.

[1]E. E. Salpeter and H. A. Bethe, Phys. Rev. **84**, 1232 (1951).
[2]M. Gell-Mann and F. Low, Phys. Rev. **84**, 350 (1951).
[3]N. Nakanishi, Suppl. Prog. Theor. Phys. **43**, 1 (1969), and references therein.
[4]G. C. Wick, Phys. Rev. **96**, 1124 (1954).
[5]R. E. Cutkosky, Phys. Rev. **96**, 1135 (1954).
[6]P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).
[7]*The Theory of Action at a Distance in Relativistic Particle Dynamics. A Reprint Collection*, edited by E. H. Kerner (Gordon and Breach, New York, 1972), and references therein.
[8]*Relativistic Action at a Distance: Classical and Quantum Aspects*, edited by J. Llosa (Springer, Berlin, 1982), and references therein.
[9]I. T. Todorov, "Constraint Hamiltonian approach to relativistic point particle dynamics, I, SISSA preprint, Trieste, 1980; "Constraint Hamiltonian mechanics of directly interacting relativistic particles," Ref. 8, p. 213; and references therein.
[10]H. Sazdjian, Phys. Rev. D **33**, 3401 (1986).
[11]It is worthwhile to emphasize that the "abnormal" solutions occur in general for strong values of the coupling constants (cf. Refs. 3 and 5).
[12]A. A. Logunov and A. N. Tavkhelidze, Nuovo Cimento **29**, 380 (1963); A. A. Logunov, A. N. Tavkhelidze, I. T. Todorov, and O. A. Khrustalev, *ibid.* **30**, 134 (1963).
[13]R. Blankenbecler and R. Sugar, Phys. Rev. **142**, 1051 (1966).
[14]F. Gross, Phys. Rev. **186**, 1448 (1969); C **26**, 2203, 2226 (1982).
[15]M. H. Partovi and E. L. Lomon, Phys. Rev. D **2**, 1999 (1970).
[16]R. N. Faustov, Teor. Fiz. **3**, 240 (1970) [Theor. Math. Phys. **3**, 478 (1970)].
[17]I. T. Todorov, Phys. Rev. D **3**, 2351 (1971); "Quasipotential approach to the two-body problem in quantum field theory," in *Properties of Fundamental Interactions*, edited by A. Zichichi (Editrice Compositori, Bologna, 1973), Vol. 9, part C, p. 951.
[18]A. Klein and T.-S. H. Lee, Phys. Rev. D **10**, 4308 (1974).
[19]E. E. Salpeter, Phys. Rev. **87**, 328 (1952).
[20]K. Nishijima, Prog. Theor. Phys. **10**, 549 (1953); **12**, 279 (1954); **13**, 305 (1955); K. Nishijima and A. H. Singh, Phys. Rev. **162**, 1740 (1967).

[21] S. Mandelstam, Proc. R. Soc. London Ser. A **233**, 248 (1955).

[22] H. Sazdjian, Phys. Lett. B **156**, 381 (1985).

[23] In general the potential V_1 appearing in the wave equation of particle 1 may be different from the potential V_2 appearing in the wave equation of particle 2. It can be shown, however, that by Poincaré invariant canonical transformations the two equations can be transformed in such a way that V_1 becomes equal to V_2. See H. Sazdjian, Ann. Phys. (NY) **136**, 136 (1981), Appendix A.

[24] If V is chosen to increase faster than c^2 in the limit $c \to \infty$, then Eqs. (2.1) will describe two-particle systems not having nonrelativistic limits. Such a situation might be relevant for the study of systems analogous to those of the "abnormal" solutions of the BS equation.

[25] S. Deser, W. Gilbert, and E. C. G. Sudarshan, Phys. Rev. **115**, 731 (1959).

[26] In L. P. Horwitz and F. Rohrlich, Phys. Rev. D **31**, 932 (1985), the authors identify the quantum mechanical and the BS wave functions and find that an equivalence of the wave equations of two-particle relativistic quantum mechanics with the BS equation is only possible if the kernel of the latter is reduced to that of an on-mass-shell elastic scattering amplitude.

[27] The second expression in (2.12) is obtained after integrating with respect to k_L. Furthermore, one must always put a small negative imaginary part in front of quantities like $(m^2 - v^{T^2})^{1/2}$.

[28] It may also contain momentum operators coming from the couplings of the fields at the vertices y_1, y_2, x_1', x_2', but they will be omitted, in general, from the notations: they do not play any essential role in the subsequent calculations.

[29] See also Ref. 28 for a remark on the momentum dependence of the kernel K, concerning notation (2.13). The momentum operators p, v in K in (2.13) come from the radiative corrections of the external propagators. See the comment after Eqs. (2.11) and (2.12).

[30] This approximation does not destroy the compatibility of the two-particle wave equations (2.10), with V related to K through relation (2.24).

[31] R. Barbieri, M. Ciafaloni, and P. Menotti, Nuovo Cimento A **55**, 701 (1967).

[32] In confining interactions, the order of magnitude of $|\langle v^{T^2}\rangle|$ is given by the confinement scale, rather than by the mass terms m.

[33] For the sake of simplicity we have omitted from K the explicit notation of the momenta and of the Dirac matrices which might be present at the vertices.

[34] V. A. Rizov, H. Sazdjian, and I. T. Todorov, Ann. Phys. (NY) **165**, 59 (1985).

[35] H. Sazdjian, Phys. Rev. D **33**, 3425 (1986).

[36] M. Ciafaloni and P. Menotti, Phys. Rev. **140**, B929 (1965).

Boson algebra as a symplectic Clifford algebra

H. Bacry[a] and M. Boon[b]
Centre de Physique Théorique, Centre National de la Recherche Scientifique, Luminy, Case 907, F-13288 Marseille Cedex 9, France

(Received 16 February 1987; accepted for publication 1 April 1987)

It is well known that the n-fermion algebra is a complex Clifford algebra of dimension 2^{2n} with the orthogonal group $O(2n,C)$ as group of automorphisms. The n-boson algebra viewed similarly as a complex "symplectic Clifford" algebra is investigated. It is of infinite dimension and has the symplectic group $Sp(2n,C)$ as a group of automorphisms. Special attention is given to the case $n = 1$ and various bases. In particular, the matrix elements of the bases between harmonic oscillator states and their relation with special functions are investigated.

I. INTRODUCTION

The origin of the present work is a paper[1] of Biedenharn and Louck[1–3] written a few years ago where the authors investigated the action of the symplectic group on the symmetrized products of a, a^+, the basis elements of the one-boson algebra. Our purpose is therefore (i) to generalize the investigation of the symplectic group to the case of an n-boson algebra, (ii) in the case $n = 1$, to analyze the Biedenharn–Louck basis elements (symplectons) in the harmonic oscillator representation, and (iii) to study the same problem when the more general bases of Cahill and Glauber[4] are introduced. It is a remarkable fact that in this last connection, the three parameters of the hypergeometric function receive a physical interpretation while its argument is associated with the choice of basis.

A résumé of this work has been presented at the XIIIth International Colloquium on Group Theoretical Methods in Physics.[5]

II. THE SYMPLECTIC CLIFFORD ALGEBRA

The boson and fermion algebras can be defined in similar ways as follows.

Consider a complex vector space V of dimension $v = 2n$ endowed with a nonsingular bilinear form which is either symplectic (boson case) or symmetric (fermion case). To be more specific, let us take a "canonical" basis made of isotropic vectors, say $\{a_1,a_2,...,a_n,\ a_1^+,a_2^+,...,a_n^+\}$ and define the bilinear form, with the aid of those basis vectors, as follows:

$$(a_i,a_j) = (a_i^+,a_j^+) = 0, \quad (a_i,a_j^+) = \tfrac{1}{2}\delta_{ij}. \tag{2.1}$$

The complex boson algebra B_n may be defined as the universal algebra generated by the a_i's, the a_i^+'s, and a unit element 1 satisfying the following boson relations:

$$\begin{aligned} &a_ia_i - a_ja_i = 0, \quad a_i^+a_j^+ - a_j^+a_i^+ = 0,\\ &a_ia_j^+ - a_j^+a = \delta_{ij}1. \end{aligned} \tag{2.2B}$$

In other words, B_n can be considered as the set of all "polynomials" in a_i and a_i^+, where the variables a_i and a_i^+ are not independent but related by (2.2B).

Similarly the complex Fermion algebra F_n may be defined as the universal algebra generated by the a_i's, the a_i^+'s, and the unit element 1 obeying the relations

$$\begin{aligned} &a_ia_j + a_ja_i = a_i^+a_j^+ + a_j^+a_i^+ = 0,\\ &a_ia_j^+ + a_j^+a_i = \delta_{ij}1. \end{aligned} \tag{2.2F}$$

In fact, in neither case is the use of a basis necessary to define the algebra. We could better say that V is a complex vector space of dimension $2n$ endowed with a nonsingular symplectic (resp. symmetric) bilinear form denoted (x,y), where $x,y \in V$ and that B_n (resp. F_n) is generated by $V \oplus \mathbb{C}$ with

$$xy - yx = 2(x,y), \tag{2.3B}$$

respectively,

$$xy + yx = 2(x,y). \tag{2.3F}$$

One recognizes, in the fermion case, the definition of a Clifford algebra (for an even dimensional space).[6] Today, where Lie algebras are considered as special cases of super Lie algebras, it would be natural to consider F_n as an *orthogonal* Clifford algebra and B_n as a *symplectic* Clifford algebra.

One problem which is relevant for physics is to build a basis of B_n from a given basis of V. To our knowledge, the first "group theoretical basis" construction was the one given by Biedenharn and Louck[1] for B_1. The basis elements of B_1 were given the name of characteristic polynomials by these authors. Although the space V is symplectic, these characteristic polynomials were made orthonormal in the Hilbert sense. We will discuss the point later on.

In order to make clear how we can find a basis for B_n from a basis of V, we will examine the case of F_n or, more precisely, the case of the Dirac algebra (F_2). Usually, one takes a Lorentz basis for V, namely $\gamma_0, \gamma_1, \gamma_2, \gamma_3$. The Dirac algebra is 16 dimensional and the usual basis of it is given by $1, \gamma_\mu, \sigma_{\mu\nu} = (i/2)(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu), \gamma_5\gamma_\mu, \gamma_5$. These basis elements span subspaces of respective dimension 1 (scalars), 4 (V itself), 6, 4 and 1; each subspace is *invariant and irreducible* under the Lorentz group. This can be seen from the commutators with the generators of the Lorentz group $\frac{1}{2}\sigma_{\mu\nu}$ (adjoint action, i.e., the "bracket" action)

a) Permanent address: Université d'Aix Marseille II, Faculté des Sciences de Luminy.
b) Permanent address: Institut Voor Theoretische Fysica, University of Nijmegen, Netherlands.

$[\frac{1}{2}\sigma_{\mu\nu}, 1] = 0$ (scalar representation),

$$[\tfrac{1}{2}\sigma_{\mu\nu}, \gamma_\rho] = i(g_{\nu\rho}\gamma_\mu - g_{\mu\rho}\gamma_\mu) \quad \text{(vector representation } \square\text{)},$$

$$[\tfrac{1}{2}\sigma_{\mu\nu}, 1/2\sigma_{\rho\lambda}] = (i/2)(g_{\nu\rho}\sigma_{\mu\lambda} - g_{\nu\lambda}\sigma_{\mu\rho} - g_{\mu\rho}\sigma_{\nu\lambda} + g_{\mu\lambda}\sigma_{\nu\rho}) \quad (\square\square),$$

$$[\tfrac{1}{2}\sigma_{\mu\nu}, \gamma_5\gamma_\rho] = i\gamma_5(g_{\nu\rho}\gamma_\mu - g_{\mu\rho}\gamma_\nu) \quad (\square\square\square),$$

$$[\tfrac{1}{2}\sigma_{\mu\nu}, \gamma_5] = 0 \quad (\square\square\square\square). \tag{2.4}$$

Once the reduction into irreducible invariant subspaces has been made, the basis vectors of F_2 are obtained by taking *antisymmetrized* products of the basis vectors of V.

Before doing something analogous for B_n, let us make some important remarks.

Remark 1: The invariant subspaces of Eqs. (2.4) are invariant under the action of the two Casimir operators of the Lorentz group. The two Casimir operators are

$$C_1 = \tfrac{1}{2}M_{\mu\nu}M^{\mu\nu}, \tag{2.5}$$

$$C_2 = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}M^{\mu\nu}M^{\rho\sigma}, \tag{2.6}$$

where $M_{\mu\nu}$ denotes "the bracket by $\frac{1}{2}\sigma_{\mu\nu}$," namely

$$M_{\mu\nu} = \tfrac{1}{2}[\sigma_{\mu\nu}, \ ], \tag{2.7}$$

therefore

$$C_1 = \tfrac{1}{8}[\sigma_{\mu\nu}, [\sigma_{\mu\nu}, \ ]], \tag{2.8}$$

$$C_2 = \tfrac{1}{8}\epsilon_{\mu\nu\rho\sigma}[\sigma^{\mu\nu}, [\sigma^{\rho\sigma}, \ ]]. \tag{2.9}$$

It is easy to check, for instance, that

$$C_1 1 = C_2 1 = 0, \tag{2.10}$$

$$C_1\gamma_\rho = 3\gamma_\rho, \quad C_2\gamma_\rho = 0, \tag{2.11}$$

relations that mean that the Lorentz scalars are characterized by the eigenvalues (0,0) and C_1 and C_2 and the Lorentz vectors[7] by the eigenvalues (3,0).

Remark 2: Rather than the Lorentz basis for V, we could choose the already mentioned isotropic basis $a_1^\pm = \frac{1}{2}(\gamma_1 \pm i\gamma_2)$, $a_2^\pm = \frac{1}{2}(\gamma_0 \pm \gamma_3)$. The reason for that is a group theoretical one: these vectors are eigenvectors of the two operators M_{12} and M_{03} (the Cartan subalgebra of the Lorentz Lie algebra). In fact,

$$\begin{aligned} &[\tfrac{1}{2}\sigma_{12}, a_1^\pm] = \pm a_1^\pm, \quad [\tfrac{1}{2}\sigma_{03}, a_1^\pm] = 0, \\ &[\tfrac{1}{2}\sigma_{12}, a_2^\pm] = 0, \quad [\tfrac{1}{2}\sigma_{03}, a_2^\pm] = \pm a_2^\pm. \end{aligned} \tag{2.12}$$

Then, by taking all antisymmetric products of the vectors $a_1^\pm$ and $a_2^\pm$, we get a basis of F_2 which is made of eigenvectors of M_{12} and M_{03} ("canonical basis"). Up to a normalization factor, they are σ_{12}, σ_{03}, $\sigma_{01} \pm i\sigma_{02} + \sigma_{31} \mp i\sigma_{23}$, $\sigma_{01} \pm i\sigma_{02} - \sigma_{31} \pm i\sigma_{23}$, etc.

Remark 3: It is perhaps important to emphasize the fact that in this part of the paper we are interested in algebras *per se* and not in their representations. For instance, the γ's are not matrices.

Remark 4: The above decomposition of F_2 into irreducible invariant subspaces remains valid for the complex extension $O(4,C)$ of the Lorentz group. Here $O(4,C)$ is the group of automorphisms of F_2.

Let us now do something similar for B_n. For the choice of a basis for V, the things are simpler because V is symplectic and it is usual to take the canonical basis defined by (2,1). The generators of the symplectic group are the elements of order 2. A suitable basis for it is given by the following $n(n+1)/2$ generators of $\mathrm{Sp}(2n,C)$:

$$\begin{aligned} J_{ij}^- &= \tfrac{1}{4}(a_i a_j + a_j a_i), \\ J_{ij}^+ &= \tfrac{1}{4}(a_i^+ a_j^+ + a_j^+ a_i^+), \\ J_{ij} &= \tfrac{1}{4}(a_i a_j^+ + a_j^+ a_i). \end{aligned} \tag{2.13}$$

By taking symmetric tensor products of V (representation $\square$) we obtain irreducible spaces of the group $\mathrm{Sp}(2n,C)$ associated with representation $\square\square$, $\square\square\square$, etc. of dimension $\binom{2n+1}{2}$, $\binom{2n+2}{3}$, etc., respectively. Moreover, the basis vectors a_i, a_i^+ of V are eigenelements of the n operators

$$M_{ii} = [\tfrac{1}{4}(a_i a_i^+ + a_i^+ a_i), \ ]. \tag{2.14}$$

Since

$$M_{ii}a_j = -\tfrac{1}{2}\delta_{ij}a_j, \quad M_{ii}a_j^+ = \tfrac{1}{2}\delta_{ij}a_j^+. \tag{2.15}$$

In the special case of B_1, the generators of $\mathrm{Sp}(2,C)$ can be written

$$J_+ = -\tfrac{1}{2}a^{+2}, \quad J_- = \tfrac{1}{2}a^2, \quad J_3 = \tfrac{1}{4}(a^+a + aa^+) \tag{2.16}$$

as already given by Biedenharn and Louck.[1] The symmetric tensor products of a and a^+ were denoted by these authors as $\mathscr{P}_j^m$; they satisfy the following commutation relations:

$$[J_3, \mathscr{P}_j^m] = m\mathscr{P}_j^m, \tag{2.17}$$

$$[J_3, [J_3, \mathscr{P}_j^m]] + \tfrac{1}{2}[J_+, [J_-, \mathscr{P}_j^m]] + \tfrac{1}{2}[J_-, [J_+, \mathscr{P}_j^m]] = j(j+1)\mathscr{P}_j^m, \tag{2.18}$$

where the notation is familiar to every physicist [$\mathrm{Sp}(2,C)$ is isomorphic to $\mathrm{SL}(2,C)$]. Equation (2.17) is the counterpart of Eqs. (2.4) of F_2. Equation (2.18) is the counterpart of Eqs. (2.10) and (2.11). It was shown by Biedenharn and Louck that the $\mathscr{P}_j^m$ form a natural basis for B_1. We intend to generalize that result to B_n in the next section.

III. SYMMETRIC BASIS OF B_n INDUCED BY A BASIS OF V

Let $e_1, e_2, \ldots, e_\nu$ ($\nu = 2n$) be an arbitrary basis of V. Since V is symplectic, every vector is isotropic but we are not necessarily taking a canonical basis. We define a basis $\{E^k_{k_1,k_2,\ldots,k_\nu}\}$ for B_n, where k and k_i are non-negative integers satisfying

$$k_1 + k_2 + \cdots + k_\nu = k, \tag{3.1}$$

with the aid of the formula

$$\left(\sum_{i=1}^{\nu} \xi^i e_i\right)^k = \sum_{[k]} \frac{k!}{[k]!}[\xi]^{[k]} E^k_{[k]}. \tag{3.2}$$

Here we use a "multiple index" notation, i.e.,

$\sum_{[k]}$ means $\sum_{k_i=0}^{k}$ with (3.1),

$[k]!$ means $k_1!k_2!\cdots k_\nu!$,

$[\xi]^{[k]}$ means $\xi_1^{k_1}\xi_2^{k_2}\cdots\xi_\nu^{k_\nu}$,

$E^k_{[k]}$ means $E^k_{k_1k_2\cdots k_\nu}$.

It is clear that the $E^k_{[k]}$ for a given k span an irreducible representation space $B_n(k)$ for the group $\mathrm{Sp}(2n,C)$. It is the representation associated with the Young diagram $\square\square\cdots\square$

made of k boxes, a representation of dimension $(\nu+k-1)!/(k-\nu)!$. We have

$$B_n = \bigoplus_{k=0}^{\infty} B_n(k), \tag{3.3}$$

where $B_n(0)=\mathbb{C}$ and $B_n(1)=V$.

Graduation: From what is well known for the symplectic group (also orthogonal groups), it is easy to see that we have a natural graduation in B_n. In fact, let us define

$$\begin{aligned} B_n^+ &= \bigoplus_{k=0}^{\infty} B_n(2k), \\ B_n^- &= \bigoplus_{k=0}^{\infty} B_n(2k+1). \end{aligned} \tag{3.4}$$

We get

$$B_n = B_n^+ \oplus B_n^- \tag{3.5}$$

and the following graduations:

$$\begin{aligned} B_n^+ B_n^+ &= B_n^- B_n^- = B_n^+, \\ B_n^+ B_n^- &= B_n^- B_n^+ = B_n^-. \end{aligned} \tag{3.6}$$

Scalar component: According to (3.3), each element of B_n can be considered *in a unique way* as the sums of elements belonging to the $B_n(k)$. The component in the $B_n(0)$space will be called the *scalar component*. Clearly it is just a number [invariant under the group $\mathrm{Sp}(2n,C)$].

Conjugation: There exists a "conjugation" in B_n which is an antilinear automorphism C such that $C^4=1$. To show this, take a canonical basis and define the antilinear mapping $C(\lambda)$

$$a_i \to -\lambda a_i^+, \quad a_i^+ \to \lambda^{-1} a_i. \tag{3.7a}$$

Clearly it preserves the symplectic product. We have

$$\begin{aligned} C(\lambda)^4 a_i &= (\lambda^* \lambda^{-1})^2 a_i, \\ C(\lambda)^4 a_i^+ &= (\lambda \lambda^{*-1})^2 a_i^+. \end{aligned} \tag{3.7b}$$

The condition $C(\lambda)^4=1$ imposes $\lambda = \pm\rho, \pm i\rho$ with $\rho>0$. Clearly, C can be extended to the whole space B_n.

Scalar product in B_n: Let X, Y be two elements of B_n belonging to the subspaces $B_n(k)$ and $B_n(k')$, respectively, i.e., corresponding to irreducible representations ▭▭▭▭▭ (k squares) and ▭▭▭▭▭ (k' squares), respectively. When we multiply these two representations, the product contains the scalar representation if and only if $k=k'$. Therefore the products XY and YX have a zero scalar component unless $k=k'$. This scalar component will be denoted by (X,Y):

$$(X,Y) = \text{scalar component of } XY. \tag{3.8}$$

It is a symmetric scalar product in B_n^+ and an antisymmetric one in B_n^-. In other words, B_n^+ (resp. B_n^-) is a carrier space for a symplectic (resp. orthogonal) representation of $\mathrm{Sp}(2n,C)$.

Hermitian product in B_n: It is well known that the irreducible representations of $\mathrm{Sp}(2n,C)$ are irreducible with respect to its maximal compact subgroup $\mathrm{USp}(2n)$. Since all representations of this group are unitary, we must be able to endow each $B_n(k)$ with an invariant Hilbert space structure. This is obtained by defining the Hermitian products for the canonical basis elements,

$$\langle a_i|a_j\rangle = 0, \quad \langle a_i^+|a_j^+\rangle = 0, \quad \langle a_i|a_j^+\rangle = \delta_{ij}, \tag{3.9}$$

or, more generally, by making use of the conjugation C:

$$\langle x|y\rangle = (Cx,y) \quad \text{for } x,y \in V, \tag{3.10}$$

$$\langle x|y\rangle = (Cx,y) \quad \text{for } x,y \in B_n. \tag{3.11}$$

Structure constants of the algebra B_n: They are defined by

$$E^k_{[k]} E^{k'}_{[k']} = \sum_{[K]} D^{k'kK}_{[k'][k][K]} E^K_{[K]}. \tag{3.12}$$

To find the coefficients D (the structure constants) explicitly, it is simpler to start with the canonical basis a_i, a_j^+ of V. We will make use of (i) the Weyl identity for $x,y\in V$

$$e^x e^y = e^{1/2(x,y)} e^{x+y} \tag{3.13}$$

(the exponentials are not elements of B_n; they will be used as generating functions); (ii) the generalized Leibnitz formula (with global indices)

$$\frac{\partial^k}{\partial[x]^{[k]}}(fg) = \sum_{[p]=0}^{[k]} \binom{[k]}{[p]} \frac{\partial^p f}{\partial[x]^{[p]}} \frac{\partial^{k-p} g}{\partial[x]^{[k-p]}}, \tag{3.14a}$$

where

$$\binom{[k]}{[p]} = \frac{[k]!}{[p]![k-p]!} \tag{3.14b}$$

is the generalized binomial coefficient; and (iii) the definition of the $E^k_{[k]}$'s

$$E^k_{[k]} = \frac{\partial^k}{\partial[x]^{[k]}}(e^x)\bigg|_{x=0}, \tag{3.15}$$

where

$$x = \xi_i a_i + \eta_i a_i^+, \quad y = \xi'_i a_i + \eta'_i a_i^+, \tag{3.16}$$

with summation $i=1,\dots,n$ understood. It is convenient to introduce a more explicit notation for the basis (3.15):

$$E^k_{[k_1][k_2]} = \frac{\partial^k}{\partial[\xi]^{[k_1]}[\eta]^{[k_2]}}(e^{\xi_i a_i + \eta_i a_i^+})\bigg|_{\xi_i=\eta_i=0}. \tag{3.17}$$

The Weyl formula now reads

$$e^{\xi_i a_i + \eta_i a_i^+} e^{\xi'_i a_i + \eta'_i a_i^+} = e^{(1/2)(\xi_i\eta'_i - \eta_i\xi'_i)} e^{(\xi_i+\xi'_i)a_i + (\eta_i+\eta'_i)a_i^+} \tag{3.18}$$

and the Leibnitz formula becomes

$$\begin{aligned} &\frac{\partial^{k+k'}}{\partial[\xi]^{[k_1]}\partial[\eta]^{[k_2]}\partial[\xi']^{[k'_1]}\partial[\eta']^{[k'_2]}}(fg) \\ &= \sum_{[p_1]=0}^{[k_1]} \sum_{[p_2]=0}^{[k_2]} \sum_{[p'_1]=0}^{[k'_1]} \sum_{[p'_2]=0}^{[k'_2]} \binom{[k_1]}{[p_1]}\binom{[k_2]}{[p_2]}\binom{[k'_1]}{[p'_1]}\binom{[k'_2]}{[p'_2]} \\ &\times \frac{\partial^{p+p'} f}{\partial[\xi]^{[p_1]}\partial[\eta]^{[p_2]}\partial[\xi']^{[p'_1]}\partial[\eta']^{[p'_2]}} \frac{\partial^{k+k'-p-p'} g}{\partial[\xi]^{[k_1-p_1]}\partial[\eta]^{[k_2-p_2]}\partial[\xi']^{[k'_1-p'_1]}\partial[\eta']^{[k'_2-p'_2]}}, \end{aligned} \tag{3.19}$$

where $k = k_1 + k_2, p = p_1 + p_2$, etc.

Now we apply the operator

$$\frac{\partial^{k+k'}}{\partial[\xi]^{[k_1]}\partial[\eta]^{[k_2]}\partial[\xi']^{[k_1']}\partial[\eta']^{[k_2']}}$$

to both sides of Eq. (3.18) and take the value at $x = y = 0$. On the left-hand side, we get $E^{k}_{[k_1][k_2]} E^{k'}_{[k_1'][k_2']}$. On the right-hand side we have to use the Leibnitz formula for

$$f = e^{(1/2)(\xi_i\eta_i' - \eta_i\xi_i')}, \quad g = e^{(\xi_i+\xi_i')a_i + (\eta_i+\eta_i')a_i^+}.$$

We get

$$\frac{\partial^{k+k'-p-p'}g}{\partial[\xi]^{[k_1-p_1]}\partial[\eta]^{[k_2-p_2]}\partial[\xi']^{[k_1'-p_1']}\partial[\eta']^{[k_2'-p_2']}}\bigg|_0$$
$$= E^{k+k'-p-p'}_{[k_1+k_1'-p_1-p_1'][k_2+k_2'-p_2-p_2']}.$$

Moreover,

$$\frac{\partial^{p+p'}f}{\partial[\xi]^{[p_1]}\partial[\eta]^{[p_2]}\partial[\xi]^{[p_1']}\partial[\eta]^{[p_2']}}
= \frac{\partial^{p}}{\partial[\xi]^{[p_1]}\partial[\eta]^{[p_2]}}
\times\left[\frac{(-)^{p_1'}}{2^{p'}}[\xi]^{p_1'}[\eta]^{[p_2']}e^{(1/2)(\xi_i\eta_i' - \xi_i'\eta_i)}\right],$$

and the only terms in this derivative that do not vanish when $x = y = 0$ are those corresponding to $[p_1] = [p_2']$ and $[p_2] = [p_1']$ for which we obtain $(-)^{p_2}([p_1]![p_2]!/2^p)$. As a conclusion, we obtain the following structure constants formula:

$$E^{k}_{[k_1][k_2]}E^{k'}_{[k_1'][k_2']}
= \sum_{[p_1]=0}^{\mathrm{Inf}([k_1],[k_2'])}\sum_{[p_2]=0}^{\mathrm{Inf}([k_2],[k_1'])}\frac{(-)^{p_2}}{2^p}[p_1]![p_2]!
\times\binom{[k_1]}{[p_1]}\binom{[k_2]}{[p_2]}\binom{[k_1']}{[p_1']}\binom{[k_2']}{[p_2']}
\times E^{k+k'-2p}_{[k_1+k_1'-p][k_2+k_2'-p]}, \tag{3.20}$$

where $\mathrm{Inf}([k_1],[k_2'])$ means a global index obtained by taking for each variable the smallest corresponding exponent appearing in $[k_1]$ and $[k_2']$.

Remark 1: The maximum value of $k + k' - 2p$ is $k + k'$. The difference term $2p$ is due to the Z_2 graduation of the algebra.

Remark 2: If we reverse the order of the product and permute the dummy indices $[p_1]$ and $[p_2]$, we see that $E^{k'}_{[k_1'][k_2']}E^{k}_{[k_1][k_2]}$ will be given by Eq. (3.20) where $(-)^{p_2}$ is to be replaced by $(-)^{p_1}$.

Remark 3: We can easily check in Eq. (3.20) that to get a nonzero scalar component on the rhs, it is necessary to have $[k_1'] = [k_2]$ and $[k_2'] = [k_1]$. In such a case, we get, using (3.8),

$$(E^{k}_{[k_1][k_2]},E^{k}_{[k_2][k_1]}) = \frac{(-)^{k_1}}{2^k}[k_1]![k_2]! \tag{3.21}$$

and one can easily verify that it is a symmetric or antisymmetric scalar product depending on whether k is even or odd. In fact, from Remark 2,

$$(E^{k}_{[k_2][k_1]},E^{k}_{[k_1][k_2]}) = \frac{(-)^{k_1}}{2^k}[k_2]![k_1]!,$$

which has the same sign as (3.21) iff $k_1 + k_2$ is even and the opposite sign iff $k_1 + k_2$ is odd.

Remark 4: Let us define the conjugation C by taking $\lambda = 1$ in (3.7):

$$Ca_i = -a_i^+, \quad Ca_i^+ = a_i, \tag{3.22}$$

we easily get from Eq. (3.2)

$$\left[\sum_{i=1}^{n}(\xi_i a_i + \eta_i a_i^+)\right]^n
= \sum_{[k_1][k_2]}\frac{k_1!k_2!}{[k_1]![k_2]!}[\xi]^{[k_1]}[\eta]^{[k_2]}E^{k}_{[k_1][k_2]},$$

$$C\left[\sum_{i=1}^{n}(\xi_i a_i + \eta_i a_i^+)\right]^n
= \sum_{[k_1][k_2]}\frac{k_2!k_1!}{[k_2]![k_1]!}[\eta^*]^{[k_2]}[-\xi^*]^{[k_1]}E^{k}_{[k_2][k_1]}$$

(summation over i is made explicit here for more clarity).

Therefore

$$CE^{k}_{[k_1][k_2]} = (-)^{k_1}E^{k}_{[k_2][k_1]}, \tag{3.23}$$

and the Hermitian product for two basis elements of B_n is given by

$$\langle E^{k}_{[k_1][k_2]}|E^{k'}_{[k_1'][k_2']}\rangle = (-)^{k_1}(E^{k}_{[k_2][k_1]}E^{k'}_{[k_1'],[k_2']}), \tag{3.24}$$

that is $(-)^{k_1}$ times the scalar component of the product $E^{k}_{[k_2][k_1]}E^{k'}_{[k_1'][k_2']}$.

According to Remark 3, Eq. (3.24) is zero unless $k_2 = k_2'$ and $k_1 = k_1'$. Therefore

$$\langle E_{[k_1][k_2]}|E_{[k_1'][k_2']}\rangle
= ([k_1]![k_2]!/2^{k_1+k_2})\delta_{[k_1][k_1']}\delta_{[k_2][k_2']}, \tag{3.25}$$

a positive definite scalar product, as expected.

IV. OTHER BASES FOR B_n

For group theoretical considerations, it was natural to consider "symmetric polynomials" in a_i and a_i^+ as a basis for B_n. Other bases can be introduced. Among them, there is one the physicist is familiar with, namely the one associated with *normally ordered products.* From a group theoretical point of view such a basis looks rather perverse. It would lead to basis elements which do not belong to irreducible representation spaces of Sp($2n$,C). Let us examine this special basis. The "normal" basis elements can be derived from a new generating function as follows:

$$E^{k(1)}_{[k_1][k_2]} = \frac{\partial^k}{\partial[\xi]^{[k_1]}\partial[\eta]^{[k_2]}}
\times(e^{-(1/2)\xi_i\eta_i}e^{\xi a_i + \eta_i a_i^+})\bigg|_{\xi_i=\eta_i=0}. \tag{4.1}$$

[Compare (3.17) for the *symmetric basis*; the superscript (1) is explained below.] In fact the change in the generating function is quite "small" since the phase factor is invariant under the group Sp($2n$,C). Obviously, the interest of the phase factor appears when we represent a_i, a_i^+ as annihilation and creation operators. In that case, the symmetry between a_i and a_i^+ is lost. Changing the sign of the phase in $e^{-(1/2)\xi_i\eta_i}$ transforms the normal product into an *antinormal* product which does not have the same properties even if from the B_n algebraic point of view there is no fundamental difference.

More generally, we can follow Cahill and Glauber[4] to define an s basis $E^{k(s)}_{[k_1][k_2]}$ by the formula

$$E^{k(s)}_{[k_1][k_2]} = \frac{\partial^k}{\partial[\xi]^{[k_1]}\partial[\eta]^{[k_2]}} (e^{-(s/2)\xi_i\eta_i} e^{\xi_i a_i + \eta_i a_i^+})\Big|_{\xi_i=\eta_i=0}, \quad (4.2)$$

where s takes values in the complex plane. For $s=0$ we are back to the symmetric basis Eq. (3.17) and for $s=1$ we obtain the normally ordered basis (4.1).

We can calculate the structure constants with an s basis. We give only the result. We have only to replace the factor $(-)^{p_2}$ in formula (3.20) by the factor $(s+1)^{p_1}(s-1)^{p_2}$.

Remark: The Leibnitz formula permits us to express the s-basis elements in terms of the symmetric basis elements. We get

$$E^{k(s)}_{[k_1][k_2]} = \sum_{[p_1]=0}^{\mathrm{Inf}([k_1],[k_2])} \binom{[k_1]}{[p_1]}\binom{[k_2]}{[p_2]}[p_1]! \times\left(-\frac{s}{2}\right)^{p_i} E^{k_1+k_2-2p_1}_{[k_1-p_1][k_2-p_2]}. \quad (4.3)$$

It follows that the scalar component of $E^{(k)s}_{[k_1][k_2]}$ is zero except when $[p_1]$ in (4.3) can take simultaneously the values $[k_1]$ and $[k_2]$, which is only possible if $[k_1]=[k_2]$. Therefore

$$\text{scalar part of } E^{k(s)}_{[k_1][k_2]} = \left(-\frac{s}{2}\right)^{k_1}[k_1]!\delta_{[k_1][k_2]}. \quad (4.4)$$

V. HARMONIC ANALYSIS FOR B_1

The question arises for the physicist to know what a basis element $E^{k(s)}_{[k_1][k_2]}$ looks like when a_i and a_i^+ are represented in a harmonic oscillator basis. Here we give the answer for B_1 only, in two parts: (i) in the present section for $s=0$ (symmetric "order"); (ii) in the next section, for a general value of s.

If we adopt the Biedenharn–Louck notation, we define

$$\mathscr{P}^m_j = \{2^j/[(j-m)!(j+m)!]^{1/2}\}E^{2j}_{j-m,j+m}, \quad (5.1)$$

where

$$(\xi a+\eta a^+)^{2j} = \sum_{m=-j}^{j}\binom{2j}{j+m}\xi^{j-m}\eta^{j+m}E^{2j}_{j-m,j+m}. \quad (5.2)$$

Here B_1^+ (resp. B_1^-) corresponds to the integral (resp. half-integral) representations of $\mathrm{Sp}(2n,C)$. The structure constants were calculated by Biedenharn and Louck:

$$\mathscr{P}^m_j\mathscr{P}^{m'}_{j'} = \sum_{J=|j-j'|}^{j+j'} D^{j'jJ}_{m'mM}\mathscr{P}^M_J, \quad (5.3)$$

where

$$D^{j'jJ}_{m'mM} = \left[\frac{(J+j'+j+1)!}{(2J+1)(j+j'-J)!(J+j-j')!(J+j'-j)!}\right] \times C^{j'jJ}_{m'mM} \quad (5.4)$$

and $C^{j'jJ}_{m'mM}$ are the Wigner coefficients.

The normalization factor in (5.1) is the standard one. It permits us to write

$$[J_+,\mathscr{P}^m_j] = [(j+m)(j-m+1)]^{1/2}\mathscr{P}^{m+1}_j,$$
$$[J_-,\mathscr{P}^m_j] = [(j-m)(j+m+1)]^{1/2}\mathscr{P}^{m-1}_j, \quad (5.5)$$
$$[J_3,\mathscr{P}^m_j] = m\mathscr{P}^m_j.$$

In particular,

$$\mathscr{P}^1_1 = \sqrt{2}E^2_{0,2} = \sqrt{2}(a^+)^2 = -2\sqrt{2}J_+,$$
$$\mathscr{P}^0_1 = 2E^2_{1,1} = a^+a + aa^+ = 4J_3, \quad (5.6)$$
$$\mathscr{P}^{-1}_1 = \sqrt{2}E^2_{2,0} = \sqrt{2}a^2 = 2\sqrt{2}J_-.$$

A matrix element $\langle N'|E^{2j}_{j-m,j+m}|N\rangle$ is zero except if the $j+m$ creators and $j-m$ annihilators relate $|N\rangle$ to $|N'\rangle$, that is if $N' = N+j+m-(j-m) = N-2m$. If we set $j-m=l$ (an integer) and $Q=2m$ (an integer), we have to compute the matrix elements

$$\langle N+Q|E^{2l+Q}_{l,l+Q}|N\rangle. \quad (5.7)$$

We define

$$\langle N+Q|E^{2l+Q}_{l,l+Q}|N\rangle = (1/2^l)[(N+Q)!/N!]^{1/2}l!\mu_l(Q,N). \quad (5.8)$$

This definition introduces *integers* $\mu_l(Q,N)$. We will suppose that *Q is always positive*. Whenever Q is negative, the corresponding matrix element is the complex conjugate of $\langle N+|Q|\,|E^{2l'+|Q|}_{l',l'+|Q|}|N\rangle$, where $l' = l+Q = l-|Q|$ and $N' = N+Q = N-|Q|$.

In the following section, we shall state the main properties of the numbers $\mu_l(Q,N)$; the proofs of (5.10)–(5.30) are found in the Appendix. First, we note that, in terms of the Biedenharn–Louck $\mathscr{P}^m_j$, they read

$$\mu_l(Q,N) = \left[\frac{N!(l+Q)!}{2^Q l'(N+Q)!}\right]^{1/2}\langle N+Q|\mathscr{P}^{Q/2}_{l+Q/2}|N\rangle. \quad (5.9)$$

The main properties are the following ones.

(1) They are real positive integers expressed by

$$\mu_l(Q,N) = P_l^{(N-l,Q)}(3), \quad (5.10)$$

where the $P_l^{(\alpha,\beta)}(x)$ are the Jacobi polynomials. In other words (Ref. 8, p. 170),

$$\mu_l(Q,N) = 2^l\binom{N}{l}F\left(-l,\ -l-Q,\ N-l+1,\ \frac{1}{2}\right), \quad (5.11)$$

where F is the hypergeometric function.

(2) They can be computed from binomial coefficients

$$\mu_l(Q,N) = \sum_{r=0}^{\mathrm{Inf}(l,N)} 2^r\binom{N}{r}\binom{l+Q}{r+Q} \quad (5.12)$$

or

$$\mu_l(Q,N) = (-)^l \sum_{m=0}^{l} (-2)^m \binom{N+Q+m}{m}\binom{l+Q}{l-m}. \tag{5.13}$$

(3) Contour integral

$$\mu_l(Q,N) = \frac{1}{2^{N+2Q}}\frac{1}{2\pi i}\oint_3 \frac{(t-1)^N(t+1)^{l+Q}}{(t-3)^{l+1}}\,dt, \tag{5.14}$$

where the path is around 3, leaving 1 outside.

(4) Integral

$$\mu_l(Q,N) = \frac{(-)^l}{(N+Q)!}\int_0^\infty e^{-u}u^{N+Q}L_l^Q(2u)\,du, \tag{5.15}$$

where L_l^Q are Laguerre polynomials.

(5) Generating functions

$$\sum_{l=0}^{\infty} \mu_l(Q,N)t^l = \frac{(1+t)^N}{(1-t)^{N+Q+1}}, \tag{5.16}$$

$$\sum_{l=0}^{\infty} \mu_l(Q,N)\frac{(-t)^l}{(l+Q)!} = \frac{N!}{(N+Q)!}e^t L_N^Q(2t), \tag{5.17}$$

$$\sum_{Q=0}^{\infty}\sum_{N=0}^{\infty} \mu_l(Q,N)\frac{x^Q z^N}{Q!N!} = e^{x+z}L_l(-2z-x), \tag{5.18}$$

$$\sum_{Q=0}^{\infty}\sum_{l=0}^{\infty}\sum_{N=0}^{\infty} \mu_l(Q,N)\frac{x^Q y^l z^N}{Q!l!N!} = e^{x+y+z}I_0(2\sqrt{y(2z+x)}), \tag{5.19}$$

$$\sum_{Q=0}^{\infty}\sum_{l=0}^{\infty}\sum_{N=0}^{\infty} \mu_l(Q,N)x^Q y^l z^N = \frac{1-y}{(1-x-y)(1-y-z-yz)}. \tag{5.20}$$

(6) Recurrence relations: (a) Q fixed,

$$\mu_{l+1}(Q,N+1) = \mu_l(Q,N+1) + \mu_{l+1}(Q,N) + \mu_l(Q,N); \tag{5.21}$$

(b) N fixed,

$$\mu_{l+1}(Q+1,N) = \mu_l(Q+1,N) - \mu_{l+1}(Q,N); \tag{5.22}$$

(c) l fixed,

$$2\mu_l(Q+1,N) = \mu_l(Q,N+1) + \mu_l(Q,N); \tag{5.23}$$

(d) also

$$\mu_l(Q,N+1) = \mu_l(Q+1,N) + \mu_{l-1}(Q+1,N). \tag{5.24}$$

These relations are easily seen in Fig. 1. We note that $\mu_l(Q,N)$ is the number of ways one can go from the origin of the diagram to the point (l,N,Q) by elementary steps in the directions (1,0,0), (0,1,0), (1,1,0), and (0,0,1) only. The numbers $\mu_l(0,N)$ are known as Delannoy numbers.[9,10] They have the symmetric property

$$\mu_l(0,N) = \mu_N(0,l), \tag{5.25}$$

which is readily obtained from Eq. (5.20) with $x=0$

$$\sum_{l=0}^{\infty}\sum_{N=0}^{\infty} \mu_l(0,N)y^l z^N = \frac{1}{1-y-z-yz}. \tag{5.26}$$

[Note that the only property of Delannoy numbers, in relation with Jacobi polynomials, mentioned in Ref. 9 is $\mu_N(0,N) = P_N(3)$, where P_N is the Legendre polynomial of degree N. This property follows from Eq. (5.10).]

(7) Special values

$$\mu_0(Q,N) = 1, \text{ whatever } Q \text{ and } N, \tag{5.27}$$

$$\mu_1(Q,N) = 2N+Q+1, \tag{5.28}$$

$$\mu_l(Q,0) = \binom{l+Q}{Q}. \tag{5.29}$$

(8) Polynomial property:

$\mu_l(Q,N)$ is a polynomial of degree l in Q and a polynomial of degree l in N. (5.30)

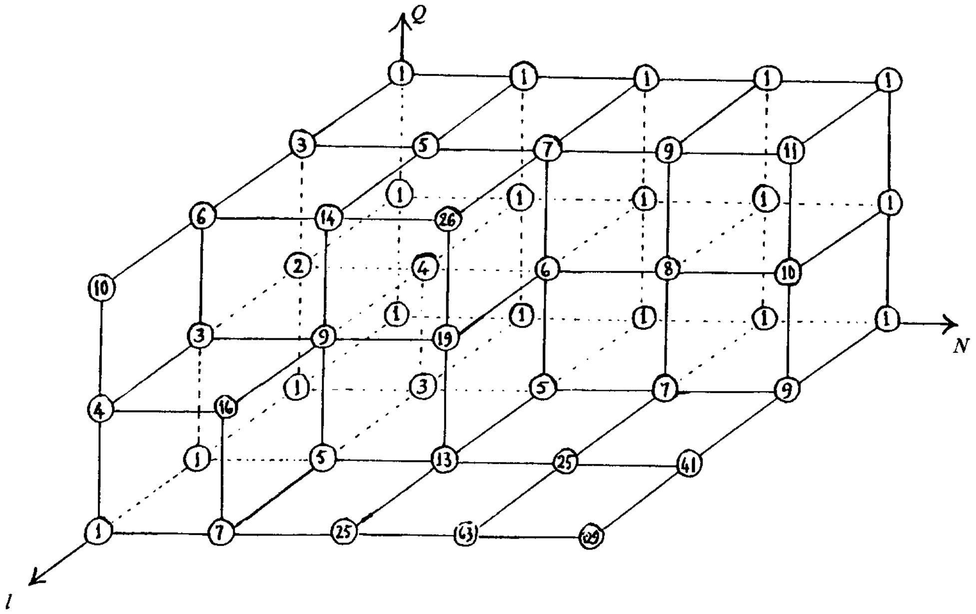

FIG. 1. The numbers $\mu_l(Q,N)$ and their recurrence relations. Relation (5.18) gives, for instance, $19 = 4 + 9 + 6$; relation (5.19), $19 = 6 + 13$; relation (5.20), $2\times 16 = 25 + 7$; relation (5.21), $25 = 19 + 6$. For $Q = 0$, we get the Delannoy numbers; for $N = 0$, we get the binomial numbers.

VI. THE CAHILL–GLAUBER BASES FOR B_1

We saw that the most natural basis for the boson algebra, from the group theoretical point of view, was the symmetric one defined in (3.17). Nevertheless it is not the one most often used in physics. Physicists prefer to use the *normal ordering*. Let us return to the more general s-ordered basis of Cahill and Glauber[4] defined in (4.2), considered for the case $n = 1$. We recall that s is any complex number. For $s = 0$, one is led to the symmetric basis and, for $s = 1$ (resp. -1), to the normal (resp. antinormal) ordering. These bases are obtained from the generating functions

$$e^{-(1/2)s\xi\eta}e^{\xi a+\eta a^{+}}=\sum_{p,q}\frac{\xi^{p}\eta^{q}}{p!q!}E^{p+q(s)}_{p,q} \tag{6.1}$$

as is easily seen from (4.2) (for $n = 1$).

It is natural to define new $\mu(Q,N)$ in the manner of (5.8) which will depend on the variable s. They will be denoted $\mu_l^{(s)}(Q,N)$ and defined by

$$\langle N+Q\,|E^{2l+Q(s)}_{l,l+Q}|N\rangle=\frac{1}{2^{l}}\sqrt{\frac{(N+Q)!}{N!}}\,\mu_l^{(s)}(Q,N), \tag{6.2}$$

a formula which generalizes (5.8).

The main properties of these functions are proved in the Appendix:

$$\mu_l^{(s)}(Q,N)=(1+s)^{l}P_l^{(N-l,Q)}\left(\frac{3-s}{1+s}\right), \tag{6.3}$$

$$\mu_l^{(s)}(Q,N)=2^{l}\binom{N}{l}F\left(-l,-l-Q,N-l+1,\frac{1-s}{2}\right), \tag{6.4}$$

$$\mu_l^{(s)}(Q,N)=\binom{l+Q}{Q}(1-s)^{l}F\left(-l,-N,Q+1,\frac{2}{1-s}\right), \tag{6.5}$$

where in (6.3) $P_l^{(\alpha,\beta)}(x)$ denotes a Jacobi polynomial of degree l.

From (6.3), it is clear that $\mu_l^{(s)}(Q,N)$ is a polynomial of degree l in s.

Other properties are

$$\mu_l^{(s)}(Q,N)=\sum_{r=0}^{\mathrm{Inf}(l,N)}2^{r}(1-s)^{l-r}\binom{N}{r}\binom{l+Q}{r+Q} \tag{6.6}$$

(they are integral numbers if s is an integer). From (6.6) we see that

$$\mu_l^{(1)}(Q,N)=2^{l}\binom{N}{l} \tag{6.7}$$

(independent of Q).

We have the recurrence relations

$$\mu^{(s)}_{l+1}(Q,N+1)=(1-s)\mu_l^{(s)}(Q,N+1)+\mu^{(s)}_{l+1}(Q,N)+(1+s)\mu_l^{(s)}(Q,N), \tag{6.8}$$

which is a consequence of (6.3) and our Jacobi recurrence relation (A24). For $s = 0$, we recover (5.21).

We have the generating functions

$$\sum_{l=0}^{\infty}\mu_l^{(s)}(Q,N)\frac{(-t)^{l}}{(l+Q)!}=\frac{N!}{(N+Q)!}e^{(s-1)t}L_n^{Q}(2t), \tag{6.9}$$

$$\sum_{l=0}^{\infty}\mu_l^{(s)}(Q,N)\,t^{l}=\frac{[1+(1+s)t]^{N}}{[1-(1-s)t]^{N+Q+1}}, \tag{6.10}$$

and the contour integral around the point $(3-s)/(1+s)$,

$$\mu_l^{(s)}(Q,N)=\frac{1}{2^{N+2Q}}\frac{(1+s)^{N+Q+l+1}}{(1-s)^{N-l}}\frac{1}{2\pi i}\times\oint_{(3-s)/(1+s)}\frac{(t-1)^{N}(t+1)^{l+Q}}{[t(1+s)-(3-s)]^{l+1}}dt \tag{6.11}$$

(the point $s = 1$ is left outside the path).

VII. SUMMARY AND CONCLUSIONS

This paper presents some results in three main areas.

(1) The analysis of the structure of an abstract n-boson algebra viewed as a complex Clifford algebra of symplectic type, as compared with the usual (orthogonal) Clifford algebra for n fermions.

(2) A more detailed study (in the above context) of the one-boson algebra, originally considered by Biedenharn and Louck,[1–3] whose basis elements are Weyl symmetrized products of the creation and annihilation operators a^{+} and a, respectively. Among other things, the matrix elements of this basis between harmonic oscillator states are related to special functions, and shown to have many interesting properties.

(3) An extension of the analysis of the one-boson algebra to a consideration of the different bases defined by a continuum of possible orderings (labeled by a complex parameter s) introduced by Cahill and Glauber.[4] These include the Weyl symmetrized ordering and normal ordering ($s = 0,1$, respectively) as special cases. The matrix elements of the basis depend here on the ordering parameter s. One of the most interesting relations with special functions is expressed in Eqs. (6.3) and (6.4). The matrix element between harmonic oscillator states $|N\rangle$ and $|N+Q\rangle$ of the s-ordered basis element constructed from l annihilation operators and $l+Q$ creation operators is seen to be essentially a hypergeometric function, or a Jacobi polynomial, wherein the three parameters and the argument are combinations of the four quantities l, Q, N, and s (the latter therefore exhaust the four "degrees of freedom" available).

Cahill and Glauber[4] are mainly interested in the properties of quantum-mechanical operators of the boson system, functions of a and a^{+}, and especially in density operators describing probability distributions on phase space. The expansion of these operators in terms of the general s-ordered basis is examined by them in detail. Such expansions are generally infinite, and Cahill and Glauber define convergence for the expansions in terms of the convergence of matrix elements (between coherent states). To correspond with this, it would be a natural next step to extend the boson algebra, considered as a symplectic Clifford algebra as in this paper, to quantities defined in terms of infinite expansions such as exponentials and other "analytic" functions. The operators $\exp(\xi a+\eta a^{+})$ used as generating functions [Eq. (6.1)] are a case in point. Another obvious extension would be to include "rational" functions, for example the Green's function $1/(\hbar\omega a^+ a - \epsilon)$. We have introduced a scalar product intrinsic to the algebra and which provides a norm independent of any representation (harmonic oscillator or other). In this context it remains to examine in what ways one can extend the algebra using convergence criteria intrinsic to the algebra, and how the class of operators obtained in such extensions correspond to the operators admitted by Cahill and Glauber and to those of interest in quantum mechanics in general.

ACKNOWLEDGMENTS

One of us (H.B.) is grateful for kind hospitality to Professor A. Janner and to the University of Nijmegen where part of this work was done.

APPENDIX: PROOF OF PROPERTIES OF THE $\mu_l(Q,N)$ AND $\mu_l^{(s)}(Q,N)$ NUMBERS

We give proofs of formulas (5.10)–(5.30) and (6.3)–(6.11).

1. Proof of formula (5.17)

From the Weyl formula

$$e^{\xi a + \eta a^+} = e^{\xi\eta/2} e^{\eta a^+} e^{\xi a}, \tag{A1}$$

we get

$$\langle N+Q | e^{\xi a + \eta a^+} | N \rangle = e^{\xi\eta/2} \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \frac{\eta^p \xi^q}{p! q!} \langle N+Q | (a^+)^p a^q | N \rangle. \tag{A2}$$

We supposed $Q \geqslant 0$. Therefore the matrix elements are zero except when $q \leqslant N$ and $p = Q + q$. Since

$$a^q |N\rangle = [N!/(N-q)!]^{1/2} |N-q\rangle,$$

$$(a^+)^{Q+q} |N\rangle = [(N+Q)!/(N-q)!]^{1/2} |N+q\rangle,$$

we get[11]

$$\langle N+Q | e^{\xi a + \eta a^+} | N \rangle = e^{(1/2)\xi\eta} \eta^Q L_N^Q(-\xi\eta) [N!/(N=Q)!]^{1/2}, \tag{A3}$$

where $L_N^Q(t)$ is the Laguerre polynomial

$$L_N^Q(t) = \sum_{q=0}^{N} \frac{(N+Q)!}{q!(Q+q)!(N-q)!} (-t)^q \tag{A4}$$

$$\left[\text{normalization: } L_N^Q(0) = \binom{N+Q}{Q}\right].$$

Also

$$\langle N+Q | e^{\xi a + \eta a^+} | N \rangle = \sum_{2j=0}^{\infty} \frac{1}{2j!} \langle N+Q | (\xi a + \eta a^+)^{2j} | N \rangle$$

$$= \sum_{2j=0}^{\infty} \sum_{m=-j}^{j} \frac{\xi^{j-m} \eta^{j+m}}{(j+m)!(j-m)!} \times \langle N+Q | E^{2j}_{j-m,j+m} | N \rangle. \tag{A5}$$

It is clear that to be nonzero the matrix element must obey the conditions $2j \geqslant Q$ and $m = Q/2$. Therefore it is natural to set

$$j = l + Q/2, \tag{A6}$$

where l is an integer. Then

$$\langle N+Q | e^{\xi a + \eta a^+} | N \rangle = \sum_{l=0}^{\infty} \frac{\xi^l \eta^{l+Q}}{(l+Q)! l!} \langle N+Q | E^{2l+Q}_{l,l+Q} | N \rangle$$

and, according to the definition (5.8) of the $\mu_l(Q,N)$:

$$\langle N+Q | e^{\xi a + \eta a^+} | N \rangle = \sum_{l=0}^{\infty} \frac{(\xi\eta)^l \eta^Q}{(l+Q)!} \frac{1}{2^l} \sqrt{\frac{(N+Q)!}{n!}} \mu_l(Q,N). \tag{A7}$$

Formula (5.17) follows by setting $\xi\eta = -t$ and by comparing (A3) and (A7):

$$\sum_{l=0}^{\infty} \mu_l(Q,N) \left(-\frac{t}{2}\right)^l \frac{1}{(l+Q)!} = \frac{N!}{(N+Q)!} e^{-t/2} L_N^Q(t). \tag{5.17}$$

In particular, for $Q = 0$

$$\sum_{l=0}^{\infty} \mu_l(0,N) \left(-\frac{t}{2}\right)^l \frac{1}{l!} = e^{-t/2} L_N(t) \tag{A8}$$

(generating function of Delannoy numbers).

2. Proof of (5.12) and (5.27)–(5.29)

Equation (5.12) is readily obtained by identification of (5.12) and (A4):

$$\sum_{l=0}^{\infty} \mu_l(Q,N) \left(-\frac{t}{2}\right)^l \frac{1}{(l+Q)!}$$

$$= \frac{N!}{(N+Q)!} \sum_{r=0}^{\infty} \left(-\frac{t}{2}\right)^r \frac{1}{r!} \times \sum_{q=0}^{N} \frac{(N+Q)!}{q!(Q+q)!(N-q)!} (-t)^q$$

$$= N! \sum_{l=0}^{\infty} \sum_{q=0}^{\mathrm{Inf}(l,N)} \frac{2^q}{q!(Q+q)!(N-q)!(l-q)!} \left(-\frac{t}{2}\right)^l.$$

We get

$$\mu_l(Q,N) = \sum_{q=0}^{\mathrm{Inf}(l,N)} 2^q \binom{N}{q} \binom{l+Q}{q+Q}, \tag{5.12}$$

which proves that they are positive integers. The proof of Eqs. (5.27)–(5.29) is a very simple exercise.

3. Proof of (5.16)

We have

$$\frac{(1+t)^N}{(1-t)^{N+Q+1}} = \frac{1}{(1-t)^{Q+1}} \left(1 + \frac{2t}{1-t}\right)^N$$

$$= \sum_{q=0}^{N} 2^q \binom{N}{q} \frac{t^q}{(1-t)^{q+Q+1}}$$

$$= \sum_{q=0}^{N} 2^q \binom{N}{q} t^q \sum_{m=0}^{\infty} \binom{q+Q+m}{q+Q} t^m.$$

Thus

$$\frac{(1+t)^N}{(1-t)^{N+Q+1}} = \sum_{l=0}^{\infty} \mu_l(Q,N) t^l \tag{5.16}$$

[where we have taken $m + q = l$ and used Eq. (5.12)].

4. Proof of recurrence relations (5.21)–(5.23)

We use three versions of formula (5.16):

$$\frac{(1+t)^N}{(1-t)^{N+Q+1}} = \sum_{l=0}^{\infty} \mu_l(Q,N)t^l, \quad (A9)$$

$$\frac{(1+t)^{N+1}}{(1-t)^{N+Q+2}} = \sum_{l=0}^{\infty} \mu_l(Q,N+1)t^l, \quad (A10)$$

$$\frac{(1+t)^N}{(1-t)^{N+Q+2}} = \sum_{l=0}^{\infty} \mu_l(Q+1,N)t^l. \quad (A11)$$

From

$$\frac{(1+t)^N}{(1-t)^{N+Q+1}} + \frac{(1+t)^{N+1}}{(1-t)^{N+Q+2}} = 2\frac{(1+t)^N}{(1-t)^{N+Q+2}} \quad (A12)$$

we get the recurrence relation for fixed l,

$$\mu_l(Q,N) + \mu_l(Q,N+1) = 2\mu_l(Q+1,N). \quad (5.23)$$

If we multiply (A9) by $1+t$ and (A10) by $1-t$, we obtain two expansions for $(1+t)^{N+1}/(1-t)^{N+Q+1}$:

$$\sum_{l=0}^{\infty} \mu_l(Q,N)t^l(1+t) = \sum_{l=0}^{\infty} \mu_l(Q+N+1)t^l(1-t)$$

from which we get the recurrence relation for fixed Q:

$$\mu_{l+1}(Q,N+1) = \mu_l(Q,N+1) + \mu_{l+1}(Q,N) + \mu_l(Q,N). \quad (5.21)$$

If we multiply (A11) by $(1-t)$ and compare with (A9) we get the recurrence relation for fixed N

$$\mu_{l+1}(Q+1,N) = \mu_l(Q+1,N) + \mu_{l+1}(Q,N). \quad (5.22)$$

Finally, we obtain formula (5.24) by multiplying (A11) by $1+t$ and comparing with (A10).

5. Proof of Eq. (5.19) and Eq. (5.18)

Let us show that the $\mu_l(Q,N)$ defined by (5.19) are the ones given in Eq. (5.12). We have from the I_0 series

$$\mu_l(Q,N) = \frac{\partial^Q}{\partial x^Q}\frac{\partial^N}{\partial z^N}\frac{\partial^l}{\partial y^l} \times \left[e^{x+y+z}\sum_{m=0}^{\infty}\frac{y^m(2z+x)^m}{(m!)^2}\right]_{x=y=z=0}$$

or

$$\mu_l(Q,N) = \frac{\partial^Q}{\partial x^Q}\frac{\partial^N}{\partial z^N} \times \left[e^{x+z}\sum_{k=0}^{l}\binom{l}{k}\frac{(2z+x)^k}{k!}\right]_{x=z=0} \quad (A13)$$

$$= \sum_{q=0}^{\mathrm{Inf}(l,N)} 2^q\binom{N}{q}\sum_{p=0}^{\mathrm{Inf}(Q,l-q)}\binom{Q}{p}\binom{l}{p+q}$$

$$= \sum_{q=0}^{\mathrm{Inf}(l,N)} 2^q\binom{N}{q}\binom{l+Q}{q+Q},$$

which proves Eq. (5.19). The proof of Eq. (5.18) is quite immediate. It is a consequence of Eq. (A13) and the property

$$\sum_{k=0}^{l}\binom{l}{k}\frac{(2z+x)^k}{k!} = L_l(-2z-x). \quad (A14)$$

Remark: If we make $z=0$ in Eq. (5.19), we get

$$e^{x+y}I_0(2\sqrt{xy}) = \sum_{Q=0}^{\infty}\sum_{l=0}^{\infty}\frac{x^Q y^l}{Q!\,l!}\mu_l(Q,0). \quad (A15)$$

This function is the generating function of binomial coefficients since, from Eq. (5.12) we have

$$\mu_l(Q,0) = \binom{l+Q}{Q} \quad (A16)$$

(the symmetry between x and y reflects the symmetry between Q and l).

6. Proof of generating functions (5.20) and (5.26)

Let us define

$$F(x,y,z) = \sum_{Q=0}^{\infty}\sum_{l=0}^{\infty}\sum_{N=0}^{\infty}\mu_l(Q,N)x^Q y^l z^N. \quad (A17)$$

If we multiply the recurrence relation (5.21) by $x^Q y^{l+1} z^{N+1}$ and sum over Q, l, N, we obtain

$$F(x,y,z) - \sum_{Q=0}^{\infty}\sum_{N=0}^{\infty}\mu_0(Q,N)x^Q z^N - \sum_{Q=0}^{\infty}\sum_{l=0}^{\infty}\mu_l(Q,0)x^Q y^l - \sum_{Q=0}^{\infty}\mu_0(Q,0)x^Q$$

$$= y\left[F(x,y,z) - \sum_{Q=0}^{\infty}\sum_{l=0}^{\infty}\mu_l(Q,0)x^Q y^l\right] + z\left[F(x,y,z) - \sum_{Q=0}^{\infty}\sum_{N=0}^{\infty}\mu_0(Q,N)x^Q z^N\right] + yzF(x,y,z)$$

or

$$F(x,y,z)[1-y-z-yz] = \frac{1}{(1-x)(1-z)} + \frac{1}{1-x-y} - \frac{1}{1-x} - \frac{y}{1-x-y} - \frac{z}{(1-x)(1-z)}$$

and

$$F(x,y,z) = \frac{1-y}{(1-x-y)(1-y-z-yz)}. \quad (5.20)$$

Here we have used the following properties:

$$\sum_{Q=0}^{\infty}\sum_{N=0}^{\infty}\mu_0(Q,N)x^Q z^N = \frac{1}{(1-x)(1-z)}, \quad (A18)$$

$$\sum_{Q=0}^{\infty}\sum_{l=0}^{\infty}\mu_l(Q,0)x^Q y^l = \frac{1}{1-x-y}, \quad (A19)$$

$$\sum_{Q=0}^{\infty}\mu_0(Q,0) = \frac{1}{1-x}. \quad (A20)$$

Equation (5.26) follows directly from Eq. (5.20) by taking $x=0$.

7. Proof of formulas (5.15), (5.13), and (5.24) and of property (5.30)

We prove formula (5.15) by considering it as a definition of $\mu_l(Q,N)$ and showing the correctness of this assumption. We see that the recurrence relation of Laguerre polynomials[8]

$$L_l^Q(2u) = L_l^{Q+1}(2u) - L_{l-1}^{Q+1}(2u)$$

provides us with the recurrence formula

$$\mu_l(Q,N+1) = \mu_l(Q+1,N) + \mu_{l-1}(Q+1,N),$$

which is our recurrence equation (5.24). Moreover, the "initial conditions" are

$$\mu_0(Q,N) = \frac{1}{(N+Q)!}\int_0^\infty e^{-u} u^{N+Q}\, du = 1,$$

$$\mu_1(Q,N) = -\frac{1}{(N+Q)!}\int_0^\infty e^{-u} u^{N+Q} L_1^Q(2u)du, \tag{A21}$$

and, since

$$L_1^Q(2u) = Q + 1 - 2u,$$

we get

$$\mu_1(Q,N) = Q + 1 + 2N. \tag{A22}$$

Equations (A21) and (A22) are exactly the ones obeyed by our original $\mu(Q,N)$ [see Eqs. (5.27) and (5.28)]. This proves Eq. (5.15) and incidentally Eq. (5.24).

Now from (5.15) and[8]

$$L_l^Q(x) = \sum_{m=0}^{l} (-)^m \binom{l+q}{l-m} \frac{x^m}{m!},$$

it is a simple matter to obtain

$$\mu_l(Q,N) = (-)^l \sum_{m=0}^{l} (-2)^m \binom{N+Q+m}{m} \binom{Q+l}{l-m}. \tag{5.13}$$

As a consequence $\mu_l(Q,N)$ is of degree l both in N and in Q [Property (5.30)].

8. Relation with Jacobi polynomials: Proof of Eqs. (5.10), (5.11), and (5.14) and of (6.3)–(6.11)

The relations of Sec. V that remain to be proved, Eqs. (5.10), (5.11), and (5.14) are special cases for $s = 0$ of Eqs. (6.3), (6.4), and (6.11), respectively, so we shall concentrate on demonstrating Eqs. (6.3)–(6.11).

The definition of $\mu_l^{(s)}(Q,N)$ from Eqs. (6.1) and (6.2) is obtained from

$$\langle N+Q | e^{-(1/2)s\xi\eta} e^{\xi a + \eta a^*} | N\rangle = \left[\frac{(N+Q)!}{N!}\right]^{1/2} \sum_{l=0}^{\infty} \frac{n^{l+Q}\xi^l}{2^l(l+Q)!} \mu_l^{(s)}(Q,N),$$

or, equivalently, using Eq. (A3) and setting $2t = -\xi\eta$, from

$$e^{(s-1)t} L_N^Q(2t) = \frac{(N+Q)!}{N!} \sum_{l=0}^{\infty} \frac{(-t)^l}{(l+Q)!} \mu_l^{(s)}(Q,N). \tag{6.9}$$

This is Eq. (6.9). The expression (6.6) for $\mu_l^{(s)}(Q,N)$ then follows by equating coefficients of t^l and using Eq. (A4).

To prove the generating relation Eq. (6.10) we write

$$\frac{[1+(1+s)u]^N}{[1-(1-s)u]^{N+Q+1}} = \frac{1}{[1-(1-s)u]^{Q+1}} \left[1 + \frac{2u}{[1-(1-s)u]}\right]^N,$$

expand the rhs in powers of u, and use Eq. (6.6) for $\mu_l^{(s)}(Q,N)$. The procedure is the same as for the $s = 0$ case in Sec. III of this appendix.

We now turn to the expressions (6.3)–(6.5) for $\mu_l^{(s)}(Q,N)$ in terms of special functions, which allow us to extend their definition to general complex values of the variables l, Q, N and s. They follow from the integral representation of the Legendre polynomial [Ref. 11, 9.211(1) and 9.215]:

$$L_N^Q(2t) = \frac{1}{2^Q\Gamma(N+1)\Gamma(-N)} \int_{-1}^{1} (1-\sigma)^{N+Q} \times (1+\sigma)^{-N-1} e^{(\sigma+1)t} d\sigma \tag{A23}$$

valid for $\mathrm{Re}(Q+1) > \mathrm{Re}(-N) > 0$. Setting this in Eq. (6.10) (above) and rearranging one finds

$$\mu_l^{(s)}(Q,N) = \frac{\Gamma(l+Q+1)}{\Gamma(l+1)} \frac{(-)^l(1+s)^l}{\Gamma(N+Q+1)\Gamma(-N)} \times \int_0^1 \tau^{N+Q}(1-\tau)^{-N-1}\left(1 - \frac{2\tau}{1+s}\right)^l d\tau. \tag{A24}$$

Thus [Ref. 11, 9.111 and 9.131(1)] we have Eq. (6.5),

$$\mu_l^{(s)}(Q,N) = \frac{\Gamma(l+Q+1)}{\Gamma(l+1)\Gamma(Q+1)} (1-s)^l \times F\left(-l, -N, Q+1, \frac{2}{1-s}\right). \tag{6.5}$$

Equation (6.4) follows directly from Eq. (6.5) using a relation between hypergeometric functions [Ref. 8, p. 170, Eq. (16)]:

$$\mu_l^{(s)}(Q,N) = 2^l \frac{\Gamma(N+1)}{\Gamma(l+1)\Gamma(N-l+1)} \times F\left(-l, -l-Q, N-l+1, \frac{1-s}{2}\right). \tag{6.4}$$

We can express the hypergeometric function in (6.5) in terms of Jacobi polynomials [Ref. 8, p. 170, Eq. (6)] to get Eq. (6.3):

$$\mu_l^{(s)}(Q,N) = (1+s)^l P_l^{(N-l,Q)}((3-s)/(1+s)). \tag{6.3}$$

We remark that, although the integral representations (A23) and (A24) are valid in the restricted range of parameters indicated under Eq. (A23), Eqs. (6.3)–(6.5) are valid for all values of the parameters where the corresponding functions are well defined; they are valid in particular for non-negative integral values of l, N, and Q. In the latter case it is easy to check that the polynomials are finite and formulas (6.3)–(6.5) agree with the expansion (6.6) for $\mu_l^{(s)}(Q,N)$.

Finally, formula (6.11) is found from the contour formula for Jacobi polynomials[8]

$$P_l^{(\alpha,\beta)}(u) = \frac{1}{2\pi i}\oint_0 \frac{(t^2-1)^l}{(t^2-u)^{l+1}} \times \left(\frac{1-t}{1+t}\right)^\alpha \left(\frac{1+t}{1+u}\right)^\beta \frac{1}{2^l} dt, \tag{A25}$$

where the contour is around 0 and leaves ± 1 outside. We set $u = 3 - s/1 + s$, $\alpha = N - l$, and $\beta = Q$ in (A25) and substitute in Eq. (6.3).

Remarks: (1) From Eqs. (6.3) and (6.5) for $\mu_l^{(s)}(Q,N)$, we see that for integral $l \geqslant 0$ it is a polynomial of degree l in the variables Q and N separately. However if N is an integer $\geqslant 0$ and $N = l$, the polynomial cuts off at degree N [cf. Eq. (6.6)].

(2) The $\mu_l^{(s)}(Q,N)$ are integral when l, N, Q are integers $\geqslant 0$ and s is integral.

(3) The recurrence relations (5.22) to (5.24) (for $s = 0$) could be obtained from well-known Jacobi polynomial recurrence relations, and could also be generalized to $s \neq 0$ by these means. Recurrence relations (5.21) (for $s = 0$) and its generalization Eq. (6.8) for $s \neq 0$ follows from the beautiful relations (not standard)

$$[(1-t)/2]P_l^{(\alpha-1,\beta)}(t) - P_{l+1}^{(\alpha-1,\beta)}(t) - P_l^{(\alpha,\beta)}(t) + P_{l+1}^{(\alpha,\beta)}(t) = 0. \quad \text{(A26)}$$

[1]L. C. Biedenharn and J. D. Louck, Ann. Phys. NY **63**, 459 (1971).
[2]L. C. Biedenharn and J. D. Louck, Lett. Math. Phys. **1**, 233 (1976).
[3]L. C. Biedenharn and J. D. Louck, *The Racah–Wigner Algebra in Quantum Theory* (Addison–Wesley, Reading, MA, 1981), pp. 243–258.
[4]K. E. Cahill and R. J. Glauber, Phys. Rev. **177**, 1857, 1882 (1969).
[5]H. Bacry and M. Boon, *Proceedings of the XIIth International Colloquium on Group Theory Methods in Physics*, Maryland (World Scientific, Singapore, 1985).

[6]If such a fermion algebra does not look like a Clifford algebra to the reader, it is due to the fact that in physics we are used to orthonormal bases in orthogonal spaces. To illustrate this point, the reader is invited to consider the γ-Dirac algebra as a typical example of a Clifford algebra and take, instead of a Lorentz basis $\gamma_0, \gamma_1, \gamma_2, \gamma_3$, the following isotropic basis $a_1^\pm = \frac{1}{2}(\gamma_1 \pm i\gamma_2)$, $a_2^\pm = \frac{1}{2}(\gamma_0 \pm \gamma_3)$. The anticommutators of these objects are $[a_1^-, a_1^+] = [a_2^-, a_2^+] = 1$ (all others vanish). The algebra obtained this way is F_2.
[7]The reader could also check, for instance, that $\frac{1}{2}(-ix_\mu \partial_\nu + ix_\nu \partial_\mu) \times (-ix^\mu \partial_\nu + ix^\nu \partial^\mu)x^\rho = 3x^\rho$.

[8]A. Erdelyi, W. Magnus, F. Oberhettinger, and G. F. Triconi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953). Vol. 2, p. 168.
[9]A. Comtet, *Advanced Combinatorics* (Reidel, Dordrecht, 1974).
[10]R. G. Stanton and D. D. Cowan, SIAM Rev. **12**, 277 (1970).
[11]A. N. Perelomov, Sov. Phys. Usp. **20**, 703 (1977) [Usp. Fiz. Nauk. **123**, 23 (1971)].
[12]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals* (Academic, New York, 1965), p. 1037.

Factorizations of vector operators for the isotropic harmonic oscillator in an angular momentum basis

O. L. de Lange
Physics Department, University of Natal, P. O. Box 375, Pietermaritzburg 3200, South Africa

(Received 3 March 1987; accepted for publication 24 June 1987)

The factorization of four vector operators, $\mathbf{D}^{\pm}(\omega)$ and $\mathbf{D}^{\pm}(-\omega)$, which occur in a representation-independent, spectrum-generating algebra for the three-dimensional, isotropic harmonic oscillator in an angular momentum basis, is considered (ω is the angular frequency of the oscillator). The $\mathbf{D}^{\pm}(\omega)$ are quantum-mechanical analogs of the classical vectors $(1 \mp i\hat{\mathbf{L}}\times)\mathbf{F}_c(\omega)$, where $\mathbf{F}_c(\omega) = -M\omega\mathbf{r}\times\mathbf{L} + \mathbf{p}L$ is constant in a frame rotating with angular velocity $\omega\hat{\mathbf{L}}$. It is shown that these four vector operators can be factorized in two different ways to yield operators that, apart from their dependence on a constant of the motion ($\mathbf{L}^2$), are linear in either $\mathbf{p}$ or $\mathbf{r}$. In this way 20 abstract operators are obtained. The properties of these operators are discussed: (i) Twelve are ladder operators for the quantum numbers l, and l and m, in the eigenkets $|lm\rangle$ of $\mathbf{L}^2$ and L_z. In linearized, differential form six of these operators are ladder operators for the spherical harmonics in the coordinate representation, while the other six are the corresponding operators in the momentum representation. (ii) The remaining eight operators factorize linear combinations of the Hamiltonian and the dimension operator. In linearized, differential form four of these operators are ladder operators for energy and angular momentum in the radial part of the coordinate-space wave functions, while the other four are the corresponding operators in the momentum representation.

I. INTRODUCTION

Recently, de Lange and Raab[1] considered the factorization of two vector constants of the motion $\mathbf{C}^{\pm}$, which occur in a representation-independent, invariance algebra for the Coulomb problem in an angular momentum basis. The $\mathbf{C}^{\pm}$ are quantum-mechanical analogs of the classical, conserved vectors

$$\mathbf{C}_c^{\pm} = (1 \mp i\hat{\mathbf{L}}\times)\mathbf{A}_c, \tag{1}$$

where $\mathbf{A}_c$ is the Laplace–Runge–Lenz vector[2] and $\mathbf{L} = \mathbf{r}\times\mathbf{p}$ is the orbital angular momentum vector. The vector operators $\mathbf{C}^{\pm}$ can be factorized in two different ways to yield a rich "substructure" of 16 operators. Apart from their dependence on the constants of the motion, $\mathbf{L}^2$ and H, these 16 operators are linear in either $\mathbf{p}$ or $\mathbf{r}$. They can be linearized by replacing constants of the motion with their eigenvalues. In differential form, these linear operators are related to those obtained by the factorization method for solving the Sturm–Liouville equation.[1]

It is natural to inquire whether a similar analysis can be carried out for the three-dimensional, isotropic harmonic oscillator (hereafter referred to as the oscillator) in an angular momentum basis. Because the classical Laplace–Runge–Lenz vector for the oscillator does not generalize to a quantum-mechanical operator,[3] the treatment cannot be completely analogous. However, Bracken and Leemon[4] have recently derived vector operators as part of a representation-independent, spectrum-generating algebra for the oscillator in an angular momentum basis. In this paper we consider vector operators $\mathbf{D}^{\pm}$ closely related to those derived in Ref. 4. These operators are quantum-mechanical analogs of the classical, time-dependent vectors

$$\mathbf{D}_c^{\pm}(\omega) = (1 \mp i\hat{\mathbf{L}}\times)\mathbf{F}_c(\omega) \tag{2}$$

and $\mathbf{D}_c^{\pm}(-\omega)$, where ω is the angular frequency of the oscillator and $\mathbf{F}_c(\omega)$ is constant in a frame rotating with angular velocity $\omega\hat{\mathbf{L}}$. (See Sec. II.) Apart from their dependence on a constant of the motion ($\mathbf{L}^2$), $\mathbf{D}^{\pm}$ are quadratic functions of $\mathbf{r}$ and $\mathbf{p}$. [See Eqs. (4)–(7).]

The purpose of this paper is twofold. First, we show that by suitable factorization of $\mathbf{D}^{\pm}$ we obtain operators that, aside from their dependence on $\mathbf{L}^2$, are linear functions of either $\mathbf{p}$ or $\mathbf{r}$. Second, we study the properties of the operators derived by factorization. The factorizations of the four vector operators $\mathbf{D}^{\pm}(\omega)$ and $\mathbf{D}^{\pm}(-\omega)$ are presented in Secs. III and IV; they yield a total of 20 operators. The properties of these abstract operators are discussed in Sec. V. Twelve of them are the same as those obtained in Ref. 1 by factorizing the operators $\mathbf{C}^{\pm}$ for the Coulomb problem: they are ladder operators for the quantum numbers l, and l and m, in the eigenkets $|lm\rangle$ of $\mathbf{L}^2$ and L_z [Eq. (32)]. After linearization (that is, replacing $\mathbf{L}^2$ with its eigenvalues), six of them are linear in $\mathbf{p}$, the other six are linear in $\mathbf{r}$.

Four of the remaining eight operators are linear in $\mathbf{p}$, apart from their dependence on $\mathbf{L}^2$. They factorize operators that are linear combinations of the Hamiltonian and the dimension operator [Eq. (36)]. Linearization yields operators that factorize a radial Hamiltonian for the oscillator [Eqs. (40) and (41)] and that are ladder operators for the energy and angular momentum in the eigenkets of this radial Hamiltonian [Eqs. (43) and (45)]. The corresponding differential forms are ladder operators for the radial part of the coordinate-space wave functions.

Finally, there are four operators that, apart from their dependence on $\mathbf{L}^2$, are linear in $\mathbf{r}$. Their properties are similar to those of the four operators discussed in the previous paragraph. In linear, differential form they are ladder operators

 0022-2488/87/112650-05$02.50

for the radial part of the momentum-space wave functions [Eqs. (53) and (54)].

II. DERIVATION AND PROPERTIES OF THE OPERATORS $\mathbf{D}^{\pm}$

We summarize the properties, and give a brief derivation, of the operators $\mathbf{D}^{\pm}$ considered in this paper.

Bracken and Leemon[4] have presented an algebraic, representation-independent solution for the oscillator in an angular momentum basis. In this basis the set of commuting observables is

$$H=(2M)^{-1}\mathbf{p}^2+\tfrac{1}{2}M\omega^2\mathbf{r}^2, \tag{3}$$

$\mathbf{L}^2$ and L_z, and the normalized, common eigenvectors are denoted by $|Elm\rangle$. The solution given in Ref. 4 is based on the vector operators

$$\mathbf{D}^{\pm}(\omega)=\mathbf{F}^{\pm}(\omega)\pm i\mathbf{G}^{\pm}(\omega), \tag{4}$$

where

$$\mathbf{F}^{\pm}(\omega)=-M\omega\mathbf{r}\times\mathbf{L}+\hbar\mathbf{p}(S\pm\tfrac{1}{2}), \tag{5}$$

$$\mathbf{G}^{\pm}(\omega)=\mathbf{p}\times\mathbf{L}+M\hbar\omega\mathbf{r}(S\pm\tfrac{1}{2}), \tag{6}$$

and

$$S=(\hbar^{-2}\mathbf{L}^2+\tfrac{1}{4})^{1/2}. \tag{7}$$

Here S is equal to one-half the dimension operator[5]; it is a Hermitian, integral operator that satisfies the eigenvalue equation

$$S|Elm\rangle=(l+\tfrac{1}{2})|Elm\rangle\,. \tag{8}$$

The operators $\mathbf{D}^{\pm}$ differ from those in Ref. 4 by a factor that depends on the constants of the motion H and $\mathbf{L}^2$; for our purposes, this difference is unimportant.[6] In what follows we suppose that $\omega>0$.

The four vector operators $\mathbf{D}^{\pm}(\omega)$ and $\mathbf{D}^{\pm}(-\omega)$ yield the 12 shift operations[7]

$$D_k^{\pm}(\omega)|Elm\rangle=\alpha_k^{\pm}(\omega)|E\pm h\omega,l\pm 1,m+k\rangle \tag{9}$$

and

$$D_k^{\pm}(-\omega)|Elm\rangle=\alpha_k^{\pm}(-\omega)|E\mp h\omega,l\pm 1,m+k\rangle, \tag{10}$$

where $k=\pm 1, 0$ and $D_{\pm 1}=D_x\pm iD_y$, $D_0=D_z$. The coefficients $\alpha_k^{\pm}$ in Eqs. (9) and (10) are functions of M, ω, and the eigenvalues E, l, and m. [See Eq. (55).] Bracken and Leemon[4] based their derivation of shift operators on properties of the boson annihilation and creation operators, and the dimension operator. In the classical limit these shift operators reduce to vectors which are (i) orthogonal to $\mathbf{L}$, and (ii) constant in a frame rotating with angular velocity $\omega\hat{\mathbf{L}}$.[4] In the following we start with the appropriate classical vectors and derive the operators $\mathbf{D}^{\pm}(\omega)$ and $\mathbf{D}^{\pm}(-\omega)$ as quantum-mechanical analogs that yield the shift operations (9) and (10). This is the same method of derivation that has been used for the Coulomb problem in an angular momentum basis.[8,9]

For the classical oscillator it is straightforward to write down two orthogonal vectors with the properties (i) and (ii) above:

$$\mathbf{F}_c(\omega)=-M\omega\mathbf{r}\times\mathbf{L}+\mathbf{p}L, \tag{11}$$

and

$$\mathbf{G}_c(\omega)=\mathbf{F}_c(\omega)\times\hat{\mathbf{L}}=\mathbf{p}\times\mathbf{L}+M\omega\mathbf{r}L. \tag{12}$$

One can readily verify that

$$\dot{\mathbf{X}}_c(\pm\omega)=\pm\boldsymbol{\omega}\times\mathbf{X}_c(\pm\omega)$$

and

$$\mathbf{X}_c^2(\pm\omega)=2M(H\pm\omega L)\mathbf{L}^2,$$

where $X=F$ or G, and $\boldsymbol{\omega}=\omega\hat{\mathbf{L}}$.

From the vectors (11) and (12) we construct the operators

$$\mathbf{F}(\omega)=-M\omega\mathbf{r}\times\mathbf{L}+\mathbf{p}K, \tag{13}$$

$$\mathbf{G}(\omega)=\mathbf{p}\times\mathbf{L}+M\omega\mathbf{r}K, \tag{14}$$

where K is a scalar constant of the motion ($[L_i,K]=[H,K]=0$) such that

$$K\to(\mathbf{L}^2)^{1/2} \tag{15}$$

as $\hbar\to 0$. [Because $\mathbf{L}$ does not commute with $\mathbf{r}$ and $\mathbf{p}$, Eqs. (13) and (14) are not unique quantum-mechanical analogs of Eqs. (11) and (12). However, Eqs. (13) and (14) are sufficient for our purposes.] From Eqs. (3), (13), and (14), and the canonical commutation relations for $\mathbf{r}$ and $\mathbf{p}$,

$$[H,\mathbf{F}(\omega)]=i\hbar\omega\mathbf{G}(\omega)$$

and

$$[H,\mathbf{G}(\omega)]=-i\hbar\omega\mathbf{F}(\omega).$$

Thus

$$[H,\mathbf{F}(\omega)\pm i\mathbf{G}(\omega)]=\pm\hbar\omega\{\mathbf{F}(\omega)\pm i\mathbf{G}(\omega)\}, \tag{16}$$

so that $\mathbf{F}\pm i\mathbf{G}$ are energy shift operators. Similarly,

$$[L_i,F_j(\omega)\pm iG_j(\omega)]=i\hbar\epsilon_{ijk}\{F_k(\omega)\pm iG_k(\omega)\} \tag{17}$$

and

$$[\mathbf{L}^2,\mathbf{F}(\omega)\pm i\mathbf{G}(\omega)]=\pm 2\hbar\{\mathbf{F}(\omega)\pm i\mathbf{G}(\omega)\}K-2i\hbar(M\omega\mathbf{r}\mp i\mathbf{p})(K^2\mp\hbar K-\mathbf{L}^2). \tag{18}$$

Clearly, if $\mathbf{F}(\omega)\pm i\mathbf{G}(\omega)$ are to be shift operators for the quantum number l, it is necessary that the second term on the right-hand side of Eq. (18) be zero. That is,

$$K^2\mp\hbar K-\mathbf{L}^2=0. \tag{19}$$

The four roots of Eq. (19) are $\hbar(\pm S+\tfrac{1}{2})$ and $\hbar(\mp S-\tfrac{1}{2})$, where S is given by Eq. (7). The two roots that have the classical limit (15) are

$$K^{\pm}=\hbar(S\pm\tfrac{1}{2}). \tag{20}$$

Substituting Eq. (20) in Eqs. (13) and (14), we obtain Eqs. (5) and (6). Equations (16)–(18), (20), and (8), yield the shift operations (9) and (10), where the $\mathbf{D}^{\pm}$ are given by Eqs. (4)–(7). The classical limit of Eq. (4) is Eq. (2).

III. FACTORIZATIONS YIELDING OPERATORS LINEAR IN p

The four vector operators $\mathbf{D}^{\pm}(\omega)$ and $\mathbf{D}^{\pm}(-\omega)$ defined by Eqs. (4)–(7) can be factored into the product of a vector operator and a scalar. This can be done in two different ways. We state the first of these two factorizations and then outline the proof. We have

$$\mathbf{D}^{\pm}(\omega)=\mathbf{U}^{\pm}\mathbf{R}^{\pm}(\omega) \tag{21}$$

and

$$\mathbf{D}^{\pm}(-\omega)=\mathbf{U}^{\pm}\mathbf{R}^{\pm}(-\omega), \tag{22}$$

where

$$\mathbf{U}^{\pm}=\pm ir^{-1}\mathbf{r}\times\mathbf{L}+\hbar r^{-1}\mathbf{r}(S\pm\tfrac{1}{2}), \tag{23}$$

$$R^{\pm}(\omega)=r^{-1}\mathbf{r}\cdot\mathbf{p}\pm i\hbar(S\mp\tfrac{1}{2})r^{-1}\pm iM\omega r, \tag{24}$$

$r=(\mathbf{r}^2)^{1/2}$, and S is given by Eq. (7).

To prove Eq. (21) we first write Eq. (23) as

$$\mathbf{U}^{\pm}=\pm ir^{-1}\mathbf{r}(\mathbf{r}\cdot\mathbf{p})\mp ir\mathbf{p}+\hbar r^{-1}\mathbf{r}(S\pm\tfrac{1}{2}).$$

Here S is a function of $\mathbf{L}^2$ and therefore it commutes with all the operators in $R^{\pm}(\omega)$ in Eq. (24). Multiplying out the product $\mathbf{U}^{\pm}R^{\pm}(\omega)$, then using the commutators $[\mathbf{r}\cdot\mathbf{p},r^{-1}]=i\hbar r^{-1}$, $[\mathbf{r}\cdot\mathbf{p},r]=-i\hbar r$, $[\mathbf{p},r^{-1}]=i\hbar r^{-3}\mathbf{r}$, and $[\mathbf{p},r]=-i\hbar r^{-1}\mathbf{r}$ and the identity

$$\mathbf{L}^2=\mathbf{r}^2\mathbf{p}^2-(\mathbf{r}\cdot\mathbf{p})^2+i\hbar\mathbf{r}\cdot\mathbf{p}, \tag{25}$$

we find

$$\mathbf{U}^{\pm}R^{\pm}(\omega)=\{-M\omega\mathbf{r}(\mathbf{r}\cdot\mathbf{p})+M\omega\mathbf{r}^2\mathbf{p}+\hbar\mathbf{p}(S\pm\tfrac{1}{2})\}\pm i[\mathbf{r}p^2-\mathbf{p}(\mathbf{r}\cdot\mathbf{p})+i\hbar\mathbf{p}+M\hbar\omega\mathbf{r}(S\pm\tfrac{1}{2})]. \tag{26}$$

Comparison with Eq. (5) shows that the term in curly brackets in Eq. (26) is equal to $\mathbf{F}^{\pm}(\omega)$, while comparison with Eq. (6) shows that the term in square brackets in Eq. (26) is equal to $\mathbf{G}^{\pm}(\omega)$. This proves Eq. (21); similarly for Eq. (22).

Apart from their dependence on $\mathbf{L}^2$, the ten operators $\mathbf{U}^{\pm}$, $R^{\pm}(\omega)$, and $R^{\pm}(-\omega)$ are linear in $\mathbf{p}$.

IV. FACTORIZATIONS YIELDING OPERATORS LINEAR IN r

Again we state the results and then outline the proof. We have

$$\mathbf{D}^{\pm}(\omega)=\pm i\mathbf{V}^{\pm}\mathbf{P}^{\pm}(\omega) \tag{27}$$

and

$$\mathbf{D}^{\pm}(-\omega)=\pm i\mathbf{V}^{\pm}\mathbf{P}^{\pm}(-\omega), \tag{28}$$

where

$$\mathbf{V}^{\pm}=\pm ip^{-1}\mathbf{p}\times\mathbf{L}+\hbar p^{-1}\mathbf{p}(S\pm\tfrac{1}{2}), \tag{29}$$

$$\mathbf{P}^{\pm}(\omega)=M\omega p^{-1}(\mathbf{r}\cdot\mathbf{p})\mp iM\hbar\omega(S\pm\tfrac{3}{2})p^{-1}\mp ip, \tag{30}$$

$p=(\mathbf{p}^2)^{1/2}$ and S is given by Eq. (7).

To prove Eq. (27) we first write Eq. (29) as

$$\mathbf{V}^{\pm}=\mp ip^{-1}\mathbf{p}(\mathbf{r}\cdot\mathbf{p})\pm ip\mathbf{r}+\hbar p^{-1}\mathbf{p}(S\mp\tfrac{3}{2}).$$

Here S commutes with all the operators in $P^{\pm}(\omega)$ in Eq. (30). Multiplying out the product $\pm i\mathbf{V}^{\pm}P^{\pm}(\omega)$, then using the commutators $[\mathbf{r}\cdot\mathbf{p},p^{-1}]=-i\hbar p^{-1}$, $[\mathbf{r}\cdot\mathbf{p},p]=i\hbar p$, $[\mathbf{r},p^{-1}]=-i\hbar p^{-3}\mathbf{p}$, and $[\mathbf{r},p]=i\hbar p^{-1}\mathbf{p}$ and the identity Eq. (25), we find

$$\pm i\mathbf{V}^{\pm}P^{\pm}(\omega)=-M\omega\mathbf{r}(\mathbf{r}\cdot\mathbf{p})+M\omega\mathbf{r}^2\mathbf{p}+\hbar\mathbf{p}(S\mp\tfrac{3}{2})\pm i[\mathbf{p}^2\mathbf{r}-\mathbf{p}(\mathbf{r}\cdot\mathbf{p})+M\hbar\omega\mathbf{r}(S\pm\tfrac{1}{2})]. \tag{31}$$

Hence we obtain Eq. (27); similarly for Eq. (28).

Apart from their dependence on $\mathbf{L}^2$, the ten operators $\mathbf{V}^{\pm}$, $P^{\pm}(\omega)$, and $P^{\pm}(-\omega)$ are linear in $\mathbf{r}$.

V. DISCUSSION AND INTERPRETATION

A. The operators U± and V±

The 12 operators $\mathbf{W}^{\pm}$ ($W=U$ or V) defined in Eqs. (23) and (29) are also factors of the vector operators $\mathbf{C}^{\pm}$ for the Coulomb problem in an angular momentum basis.[1] They are ladder operators for the quantum numbers l, and l and m, in the eigenkets $|lm\rangle$ of $\mathbf{L}^2$ and L_z. Specifically,[1]

$$W_k^{\pm}|lm\rangle=\hbar\beta_k^{\pm}|l\pm 1,m+k\rangle, \tag{32}$$

where $k=\pm 1,0$ and

$$|\beta_{+1}^{\pm}|^2=(l\pm m+\tfrac{1}{2}\pm\tfrac{1}{2})(l\pm m+\tfrac{1}{2}\pm\tfrac{3}{2})a_l^{\pm}, \tag{33}$$

$$|\beta_{-1}^{\pm}|^2=(l\mp m+\tfrac{1}{2}\pm\tfrac{1}{2})(l\mp m+\tfrac{1}{2}\pm\tfrac{3}{2})a_l^{\pm}, \tag{34}$$

$$|\beta_0^{\pm}|^2=(l-m+\tfrac{1}{2}\pm\tfrac{1}{2})(l+m+\tfrac{1}{2}\pm\tfrac{1}{2})a_l^{\pm}, \tag{35}$$

$$a_l^{\pm}=(2l+1)(2l+1\pm 2)^{-1}.$$

If we replace S by its eigenvalues $l\pm\frac{1}{2}$ in Eq. (23), we obtain operators linear in $\mathbf{p}$. The corresponding wave-mechanical operators are ladder operators for the quantum numbers l and m of spherical harmonics in the coordinate representation.

Similarly, from Eq. (29) we can obtain operators linear in $\mathbf{r}$, and hence wave-mechanical operators that are ladder operators for the spherical harmonics in the momentum representation. More details can be found in Ref. 1 and the references therein.

B. The operators $R^{\pm}$

It can be shown that

$$[R^{\pm}(\omega)]^{\dagger}R^{\pm}(\omega)=2MH+2M\hbar\omega(S\pm 1), \tag{36}$$

where $R^{\pm}(\omega)$ are given by Eq. (24), H by Eq. (3), S by Eq. (7), and † denotes the adjoint operator. Thus the $R^{\pm}(\omega)$ factorize the operators on the right-hand side of Eq. (36). If we replace S with its eigenvalues $l+\frac{1}{2}$ in Eq. (24) we obtain operators linear in $\mathbf{p}$, namely,

$$R_l^{\pm}(\omega)=p_r\pm i\hbar(l+\tfrac{1}{2}\pm\tfrac{1}{2})r^{-1}\pm iM\omega r. \tag{37}$$

Here

$$p_r=r^{-1}\mathbf{r}\cdot\mathbf{p}-i\hbar r^{-1} \tag{38}$$

is the canonical conjugate of r. The adjoint of Eq. (37) gives

$$[R_l^{\pm}(\omega)]^{\dagger}=R_{l\pm 1}^{\mp}(\omega). \tag{39}$$

Replacing S with $l+\frac{1}{2}$ in Eq. (36) and using Eq. (39) yields

$$R_{l\pm 1}^{\mp}(\omega)R_l^{\pm}(\omega)=2MH_l+2M\hbar\omega(l+\tfrac{1}{2}\pm 1). \tag{40}$$

Here

$$H_l=(2M)^{-1}\{p_r^2+\hbar^2 l(l+1)r^{-2}\}+M\omega^2r^2 \tag{41}$$

is the radial Hamiltonian obtained from Eq. (3) by using the identity $\mathbf{L}^2=\mathbf{r}^2\mathbf{p}^2-\mathbf{r}^2p_r^2$ to eliminate $\mathbf{p}^2$ in favor of p_r^2 and then replacing $\mathbf{L}^2$ with its eigenvalues. The factorization [Eq. (40)] has been discussed previously.[10] If $|El\rangle$ denotes an eigenket of H_l,

$$H_l|El\rangle=E|El\rangle, \tag{42}$$

it follows from Eqs. (40) and (42) that

$$R_l^{\pm}(\omega)|El\rangle = \gamma^{\pm}(\omega)|E \pm \hbar\omega, l \pm 1\rangle, \quad (43)$$

where

$$|\gamma^{\pm}(\omega)|^2 = 2M\{E + (l + \tfrac{1}{2} \pm 1)\hbar\omega\}. \quad (44)$$

The operators $R^{\pm}(-\omega)$, which appear in the factorization [Eq. (22)], are given by Eq. (24) with ω replaced by $-\omega$. For these and the corresponding operators $R_l^{\pm}(-\omega)$, results can be deduced which are similar to those given above for $R^{\pm}(\omega)$ and $R_l^{\pm}(-\omega)$ namely, Eqs. (36), (37), (39), (40), and (43), with ω replaced by $-\omega$. Thus the counterpart of Eq. (34) is

$$R_l^{\pm}(-\omega)|El\rangle = \gamma^{\pm}(-\omega)|E \mp \hbar\omega, l \pm 1\rangle. \quad (45)$$

The wave-mechanical operators obtained by substituting

$$p_r = -i\hbar r^{-1} - i\hbar\frac{\partial}{\partial r}$$

in Eq. (37) are the same as the differential operators derived using the factorization method for solving the Sturm–Liouville equation.[11] The wave-mechanical form of Eq. (45) yields a pair of first-order differential equations whose solutions are the radial coordinate-space wave functions for the oscillator.

C. The operators $P^{\pm}$

For the operators $P^{\pm}(\omega)$ defined by Eq. (30) a similar analysis to that given above for $R^{\pm}(\omega)$ can be carried out. Thus

$$[P^{\pm}(\omega)]^{\dagger}P^{\pm}(\omega) = 2MH + 2M\hbar\omega(S \pm 1). \quad (46)$$

Replacing S with $l + \frac{1}{2}$ in Eq. (30) yields the operators

$$P_l^{\pm}(\omega) = M\omega r_p \mp iM\hbar\omega(l + \tfrac{1}{2} \pm \tfrac{1}{2})p^{-1} \mp ip, \quad (47)$$

where

$$r_p = p^{-1}\mathbf{p}\cdot\mathbf{r} + i\hbar p^{-1} \quad (48)$$

is the canonical conjugate of p. The adjoint of Eq. (47) gives

$$[P_l^{\pm}(\omega)]^{\dagger} = P_{l\pm 1}^{\mp}(\omega). \quad (49)$$

From Eq. (46), with S replaced by $l + \frac{1}{2}$, and Eq. (49) we have

$$P_{l\pm 1}^{\mp}(\omega)P_l^{\pm}(\omega) = 2MH_l' + 2M\hbar\omega(l + \tfrac{1}{2} \pm 1). \quad (50)$$

Here

$$H_l' = (2M)^{-1}\mathbf{p}^2 + \tfrac{1}{2}M\omega^2\{r_p^2 + \hbar^2 l(l+1)p^{-2}\} \quad (51)$$

is the radial Hamiltonian obtained from Eq. (3) by using the identity $\mathbf{L}^2 = \mathbf{p}^2\mathbf{r}^2 - \mathbf{p}^2 r_p^2$ to eliminate $\mathbf{r}^2$ in favor of r_p^2 and then replacing $\mathbf{L}^2$ with its eigenvalues. Let $|El\rangle$ denote an eigenket of H_l',

$$H_l'|El\rangle = E|El\rangle. \quad (52)$$

[For convenience we have not distinguished the eigenkets in Eqs. (42) and (52).] It follows from Eqs. (50) and (52) that

$$P_l^{\pm}(\omega)|El\rangle = \mp i\gamma^{\pm}(\omega)|E \pm \hbar\omega, l \pm 1\rangle, \quad (53)$$

where the magnitude of the coefficient $\gamma^{\pm}(\omega)$ is given by Eq. (44). [The factor $\mp i$ has been included in Eq. (53) because of the factor $\pm i$ in Eq. (27).]

Finally, there are the operators $P^{\pm}(-\omega)$ and $P_l^{\pm}(-\omega)$ given by Eqs. (30) and (47) with ω replaced by $-\omega$. They satisfy Eqs. (46), (47), (49), (50), and (53) with ω replaced by $-\omega$. Thus in place of Eq. (53) one has

$$P_l^{\pm}(-\omega)|El\rangle = \mp i\gamma^{\pm}(-\omega)|E \mp \hbar\omega, l \pm 1\rangle. \quad (54)$$

The operators $P_l^{\pm}$ are linear in $\mathbf{r}$. Wave-mechanical operators are obtained by substituting

$$r_p = i\hbar p^{-1} + i\hbar\frac{\partial}{\partial p}$$

in Eq. (47). The solutions to the wave-mechanical form of Eq. (54) are the radial momentum-space wave functions for the oscillator.

The coefficients $\alpha_k^{\pm}(\omega)$ in Eqs. (9) and (10) can be obtained from the above results:

$$\alpha_k^{\pm}(\omega) = \hbar\beta_k^{\pm}\gamma^{\pm}(\omega), \quad (55)$$

where the magnitudes of $\beta_k^{\pm}$ and $\gamma^{\pm}(\omega)$ are given by Eqs. (33)–(35) and (44).

It is interesting to compare the factorizations presented here for the oscillator with those given previously for the Coulomb problem.[1] In each case we find the same vector operators ($\mathbf{U}$ and $\mathbf{V}$); it is the scalar operators obtained by factorization that are different. Thus for the Coulomb problem the operators analogous to $R^{\pm}(\omega)$ are[1]

$$R^{\pm} = \pm ir^{-1}\mathbf{r}\cdot\mathbf{p} - \hbar(S \mp \tfrac{1}{2})r^{-1} + \hbar a^{-1}(S \pm \tfrac{1}{2})^{-1}, \quad (56)$$

where $a = 4\pi\epsilon_0\hbar^2/(Me^2)$ is the Bohr radius. For these,[1]

$$(R^{\pm})^{\dagger}R^{\pm} = 2MH + 4\hbar^2 a^{-2}(2S \pm 1)^{-2}, \quad (57)$$

where

$$H = (2M)^{-1}\mathbf{p}^2 - \hbar^2(Ma)^{-1}r^{-1}.$$

The factorization Eq. (57) may be compared to Eq. (36). Similarly, the operators analogous to $P^{\pm}(\omega)$ are[1]

$$\begin{aligned} P^{\pm} = &\mp i\hbar^{-1}p^{-1}\mathbf{r}\cdot\mathbf{p}(\mathbf{p}^2 - 2MH) \\ &+ p^{-1}(\mathbf{p}^2 + 2MH)(S \pm \tfrac{1}{2}) \\ &\mp 2p^{-1}(\mathbf{p}^2 - 2MH). \end{aligned} \quad (58)$$

For these,[1]

$$(P^{\pm})^{\dagger}P^{\pm} = 2MH(2S \pm 1)^2 + 4\hbar^2 a^{-2}, \quad (59)$$

which may be compared with Eq. (46).

ACKNOWLEDGMENTS

I thank Professor R. E. Raab for helpful discussions.

Support from the South African Council for Scientific and Industrial Research is acknowledged.

[1] O. L. de Lange and R. E. Raab, Phys. Rev. A **35**, 951 (1987).

[2] We use the nomenclature of H. Goldstein, *Classical Mechanics* (Addison–Wesley, Reading, MA, 1980), 2nd. ed., p. 102. See also H. Goldstein, Am. J. Phys. **44**, 1123 (1976).

[3] D. M. Fradkin, Prog. Theor. Phys. **37**, 798 (1967).

[4] A. J. Bracken and H. I. Leemon, J. Math. Phys. **21**, 2170 (1980).

[5] L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics* (Addison–Wesley, Reading, MA, 1981), Chap. 6.

[6]It can be shown that $\mathbf{D}^{\pm}$ are related to the shift operators λ and ν derived in Ref. 4:

$$\lambda = i(4M\hbar^3\omega)^{-1/2}\mathbf{D}^{-}(\omega)(\hbar^{-1}\omega^{-1}H+S-1)^{-1/2}S^{-1/2},$$

$$\nu = -(2\sqrt{2}M\hbar^3\omega)^{-1}[\mathbf{D}^{+}(-\omega)\cdot\mathbf{D}^{-}(\omega)+\mathbf{D}^{-}(\omega)\cdot\mathbf{D}^{+}(-\omega)] \times(\hbar^{-1}\omega^{-1}H+S-1)^{-1/2}.$$

The corresponding relations for the adjoint operators $\lambda^{\dagger}$ and $\nu^{\dagger}$ can be determined using $[\mathbf{D}^{+}(\omega)]^{\dagger}=S\mathbf{D}^{-}(\omega)S^{-1}$.

[7]In Ref. 4, shift operations are derived for operators λ and $\lambda^{\dagger}$. These, and the relations given above in 6, yield Eqs. (9) and (10).

[8]O. L. de Lange and R. E. Raab, Phys. Lett. A **118**, 219 (1986).

[9]O. L. de Lange and R. E. Raab, Phys. Rev. A **34**, 1650 (1986).

[10]See J. D. Newmarch and R. M. Golding, Am. J. Phys. **46**, 658 (1978), and the references therein.

[11]E. Schrödinger, Proc. R. Irish Acad. A **46**, 9 (1940); **46**, 183 (1941); **47**, 53 (1941); L. Infeld and T. E. Hull, Rev. Mod. Phys. **23**, 21 (1951); P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Part I, p. 729.

Probability of convergence of perturbation theory for hydrogen photoionization

E. Caliceti and L. Goldoni
Dipartimento di Matematica Pura ed Applicata, Università di Modena, 41100 Modena, Italy
S. Graffi
Dipartimento di Matematica, Università di Bologna, 40127 Bologna, Italy

(Received 9 December 1986; accepted for publication 3 June 1987)

The probability is estimated that at a given frequency ω the action of an external, monochromatic radiation field acting on a hydrogen atom yields a fixed positive convergence radius of the Rayleigh-Schrödinger perturbation expansion for the quasienergy resonances.

I. INTRODUCTION

In this paper we consider the hydrogen atom under the action of an external, spatially homogeneous, monochromatic radiation field of strength F and frequency ω, described by the nonautonomous Schrödinger equation

$$H(t)\psi \equiv (-\Delta + V(\mathbf{r}) + Fx\cos\omega t)\psi(\mathbf{r},t) = i\,\partial_t\psi(\mathbf{r},t), \tag{1.1}$$

where $\mathbf{r} = (x,y,z)$ and $V(\mathbf{r}) = -Z/|\mathbf{r}|$ is the Coulomb potential. This system, often referred to as the AC-Stark effect, is currently under active investigation,[1] also for its relevance to the problem of "quantum chaos"[2] through the so-called "chaotic photoionization."[3]

It is convenient to analyze (1.1) using time-independent methods: in this case, employing the quasienergy, or Floquet formalism, first mathematically implemented in this kind of problems by Yajima.[4,5] Yajima's results can be extended to the case (1.1) by using the radiation gauge for the external field, which amounts to performing the unitary transformation $\psi(\mathbf{r},t) = \exp[iFx\sin\omega t/\omega]\phi(\mathbf{r},t)$. This yields

$$H'(t)\phi(\mathbf{r},t) \equiv [(-i\nabla - \omega^{-1}F\sin\omega t)^2 + V(\mathbf{r})]\phi(\mathbf{r},t) = i\,\partial_t\phi(\mathbf{r},t). \tag{1.2}$$

Since the perturbation is $2\pi/\omega$ periodic, according to Floquet theory we look for solutions of (1.2) in the form of a quasiperiodic function of t, $\phi(\mathbf{r},t) = e^{-i\lambda t}f(\mathbf{r},t)$, where $f(\mathbf{r},t)$ is an L^2-valued $2\pi/\omega$-periodic function of t. The Floquet exponent λ is given by the solution of the spectral problem

$$Kf \equiv (H'(t) - i\,\partial_t)f(\mathbf{r},t) = \lambda f(\mathbf{r},t) \tag{1.3}$$

considered in the extended Hilbert space $\mathscr{K} = L^2(\mathbb{R}^3)\otimes L^2(\mathbb{T}_\omega)$ of the $2\pi/\omega$ time-periodic space-time L^2 functions. Here $\mathbb{T}_\omega = \mathbb{R}\setminus(2\pi/\omega)$ is the circle. If indeed f is an eigenvector of K corresponding to the eigenvalue λ then (1.2) admits the solution $\psi(r,t) = e^{-i\lambda t}f(\mathbf{r},t)$ and conversely. The operator K is known also as the quasienergy operator, and its spectrum as the quasienergy spectrum.

For small values of the field strength F the spectral problem (1.1) can be analyzed by perturbation theory. The free quasienergy, or Floquet, operator is, of course,

$$K_0 := -\Delta + V(\mathbf{r}) - i\,\partial_t \tag{1.4}$$

and the perturbation is given by

$$W(F) = FW_1 + F^2W_2, \quad W_1 = 2i\omega^{-1}\sin\omega t\,\partial_x, \quad W_2 = \omega^{-2}\sin^2\omega t. \tag{1.5}$$

Deferring to Sec. II the description of the above differential expressions as operators in the Hilbert space $\mathscr{K}$, we remark here that the free quasienergy spectrum, i.e., the spectrum $\sigma(K_0)$ of K_0, is the whole of $\mathbb{R}$ and contains the doubly infinite sequence of eigenvalues $\lambda_{n,k} = -n^{-2} + k\omega = \lambda_n + k\omega$, $n = 1,2,\ldots$, $k = 0, \pm 1, \pm 2,\ldots$, embedded in the continuum. Here we have set $Z = \sqrt{2}$ to normalize the hydrogen bound states at $-n^{-2}$. In this respect the problem looks very much like the autoionization one in atoms, in which the embedded eigenvalues turn into resonances,[6] defined through dilation analyticity.[7] In fact, the analogous result has been proved for the present system.[8] Namely, the dilation analyticity technique can be implemented on the Floquet operator (1.4) to prove that all eigenvalues existing at $F = 0$ turn into resonances for $F > 0$ small.

Unlike the static field case $\omega = 0$, however (see, e.g., Ref. 9), the (Rayleigh–Schrödinger) perturbation expansion near any (simple) unperturbed quasienergy level converges to the nearby resonance.[8] Furthermore, another important difference from the static case is that the imaginary part of the resonance, which is proportional to the ionization rate, is not an exponentially small quantity in the field strength but has a power behavior. More precisely: if λ is a hydrogenic bound state, the first nonvanishing order in perturbation theory for the nearby resonance is given by the smallest integer p such that $\lambda + \omega p > 0$, i.e., by the number of photons it takes to ionize the bound state λ. Moreover, the imaginary part is expressed by Fermi's "golden rule" computed at order p.[8]

These results are, however, critically dependent on the arithmetical properties of the driving frequency ω with respect to the difference between any two hydrogenic bound states. If indeed $\lambda_{n,k} = \lambda_n + k\omega$ denote the free quasienergy levels, and if ω is such that $\lambda_n - \lambda_m = s\omega$ for some $m \neq n$ and some $s \neq 0$, then $\lambda_{n,s} = \lambda_{m,0}$, i.e., there is multiplicity doubling. The set $\{\omega_{m,n,s} = (\lambda_n - \lambda_m)/s,\ (m,n) = 1,2,\ldots,\ s = \pm 1, \pm 2,\ldots\}$ of such values of ω has measure equal to 0; however on sequences $\{\omega\}$ converging to $\omega_{m,n,s}$ the isolation distance of $\lambda_{n,s}$, i.e., its distance from any other free quasienergy level, becomes arbitrarily small. Since the convergence radius of the perturbation expansion is, roughly speaking, inversely proportional to the isolation distance, it follows that the convergence statement is of purely academic interest unless a set of values of ω can be determined for which the radius of convergence admits a lower bound independent of ω.

 0022-2488/87/112655-07$02.50

The purpose of this paper is to provide a detailed analysis of this point. In the next section we will state and comment on the results, whose proof is described in Sec. III. Finally we work out in the Appendix some details on perturbation estimates in the present non-self-adjoint case that we were unable to locate in the literature.

II. STATEMENT OF THE RESULTS

To formulate the results to be discussed and proved below, we first have to quote the relevant results of Ref. 8 on the resonances of the quasienergy operator.

Lemma: II.I: Let $0<\omega<1$, $F>0$, and θ be complex, $0<\text{Im}\,\theta<\pi/4$. Define on the Hilbert space $\mathscr{H}=L^2(\mathbb{R}^3)\otimes L^2(\mathbb{T}_\omega)$ a two-parameter operator family $K(F,\theta)$ as the action of the differential expression

$$K(F,\theta)=(-ie^{-\theta}\nabla-\omega^{-1}F\sin\omega t)^2+V(\theta)-i\frac{\partial}{\partial t}$$

on the domain $\mathscr{D}=L^2(\mathbb{T}_\omega)\otimes H^2(\mathbb{R}^3)\cap H^1(\mathbb{T}_\omega)\otimes L^2(\mathbb{R}^3)$. We abbreviate $K(0,\theta)$ by $K(\theta)$, $K(0)$ by K, and $T(0)$ by T. Here $V(\theta)=-Ze^{-\theta}/|\mathbf{r}|$, and $H^3(\cdot)$ denotes the usual Sobolev space of order s. Then the following statements hold.

(1) If $\theta\in\mathbb{R}$, $K(F,\theta)$ is essentially self-adjoint.

(2) If $\text{Im}\,\theta>0$, $K(F,\theta)$ is closed and is a type-A holomorphic family of operators (in the sense of Ref. 10, § VII.2) with respect to (F,θ) for $F\in\mathbb{C}$, $0<\text{Im}\theta<\pi/4$.

(3) *The resolvent* $R(F,\theta,z)=[K(F,\theta)-z]^{-1}$ of $K(F,\theta)$ is strongly continuous as $\text{Im}\,\theta\downarrow 0$, $\text{Im}\,z>0$.

Remarks: (a) For $\theta\in\mathbb{R}$, $F>0$ we have $\sigma(K)=\mathbb{R}$, with embedded eigenvalues $\lambda_{n,k}=-n^{-2}+k\omega$, $n\in\mathbb{N}$, $k\in\mathbb{Z}$. For $\text{Im}\,\theta>0$, $\sigma[K(\theta)]=\{\lambda_{n,k}\}\cup\{e^{-2\theta}\mathbb{R}_++s\omega,\ s\in\mathbb{Z}\}$, i.e., all embedded eigenvalues become isolated for $\text{Im}\,\theta>0$, as usual within the dilation analyticity formalism.[7]

(b) The above assertions imply that the free quasienergy levels, for $F>0$ small, turn into resonances in the standard sense of dilation analyticity. In fact, by (2) the isolated eigenvalues $\lambda_{n,k}$ of $K(\theta)$ are stable and, for $F>0$ small turn into isolated eigenvalues $\lambda_{n,k}(F)$ of $K(F,\theta)$. The eigenvalue $\lambda_{n,k}(F)$ do not depend on θ by standard arguments of dilation analyticity, and by the holomorphy near $F=0$ are given by (branches of) holomorphic functions near $F=0$. Again by the standard arguments of dilation analyticity and (1) $\text{Im}\,\lambda_{n,k}(F)<0$ [actually in this case of the Coulomb potential $\text{Im}\,\lambda_{n,k}(F)<0$ by the validity of Fermi's "golden rule"] and by (3) the complex eigenvalues $\lambda_{n,k}(F)$ are second sheet poles of the scalar products of the resolvent $R(F,\theta,z)$ taken between dilation analytic vectors. Conversely any pole of such analytic continuation of the resolvent must coincide with one of the eigenvalues $\lambda_{n,k}(F)$.

(c) The above results holds true also in the N-body case as well as in one dimension,[8] i.e., for the time-dependent Schrödinger equation on $\mathbb{R}_+(-d^2/dx^2-|x|^{-1}+Fx\cos\omega t)\psi=i\,\partial_t\psi$ with the Dirichlet boundary condition at 0. This simplified model has recently become interesting for the chaotic photoionization problem.[11]

(d) If λ_n is a simple eigenvalue of T (which is always the case in one dimension) and there is no $k\neq 0$ such that $\lambda_n-\lambda_m=k\omega$ for $n\neq m$ (which is true almost everywhere in ω) then by (2) $\lambda_{n,k}(F)$ is holomorphic near $F=0$ and the Rayleigh–Schrödinger perturbation expansion near $\lambda_{n,k}$ is therefore convergent.

The problem to be taken on here is the discussion of this last statement as a function of ω, i.e., the dependence of the radius of convergence on ω. The results are as follows.

Proposition I: Let $0<\omega<1$, and let $\omega(m,n,k)=k^{-1}(n^{-2}-m^{-2})$, $k\in\mathbb{Z}$ $(n,m)\in\mathbb{N}$ be the sequence of the resonance values of ω, i.e., those values of ω such that $n^{-2}-m^2=k\omega$ for some $m\neq n$, $k\neq 0$. Let $\lambda_{m,0}$ be a simple eigenvalue of $K(\theta)$, $0<\text{Im}\,\theta<\pi/4$, and set

$$B^{(m)}=]0,1[\setminus\bigcup_{n,k}\{\omega(m,n,k)\}. \tag{2.1}$$

Furthermore, let $r_0(m,\omega)$ denote the radius of convergence of the perturbation expansion near $\lambda_{m,0}$.

Then we have the following.

(1) $0<\omega(m,n,k)<1$, for all m,n,k.

(2) If $r_0(m,\omega)>A\geqslant 0$, for some A independent of $\omega\in B^{(m)}$, then $A=0$.

(3) Let ϵ_0 be arbitrarily fixed, and let $B^{(m)}(\epsilon_0)=\{\omega\in B^{(m)}:\ \text{dist}(\lambda_{m,0},[\sigma(K(\theta)\setminus\lambda_{m,0}])>\epsilon_0\}$. Then $B^{(m)}(\epsilon_0)\neq\varnothing$ and there is $a(\epsilon_0)>0$ independent of $\omega\in B^{(m)}(\epsilon_0)$ such that $r_0(m,\omega)>a(\epsilon_0)$.

Remark: A different wording of statement (2) is that there is no positive lower bound for $r_0(m,\omega)$ independent of $\omega\in B^{(m)}$. This motivates the search for the subset $B^{(m)}(\epsilon_0)$ for which a positive, uniform lower estimate can be obtained. A description of $B^{(m)}(\epsilon_0)$ is given in Proposition II. We will see that there is no simple expression for the isolation distance as a function of $\omega\in B^{(m)}$. Therefore it is useful, and possibly simpler, to look at the probability for a given point ω to have an isolation distance greater than ϵ_0, and this will be the content of Corollary III.

Proposition II: Let $B^{(m)}(\epsilon_0)$ be defined as in Proposition I (3), and let $\mu(\cdot)$ denote the Lebesgue measure on $\mathbb{R}$. Then we have the following.

(1) $\mu[B^{(m)}(\epsilon_0)]\to\mu[B^{(m)}]$ as $\epsilon_0\to 0$.

(2) Let $T^{(m)}(\epsilon_0)=]0,1[\setminus B^{(m)}(\epsilon_0)$, and set

$$\omega_0(\epsilon_0,m)=(\epsilon_0^{-2/3}-m^2)^{-1},\quad \epsilon_0<m^{-3}. \tag{2.2}$$

Then $]0,\omega_0(\epsilon_0,m)[\subset T^{(m)}(\epsilon_0)$.

(3) Consider the following sequences of pairwise disjoint intervals:

$$I_1(0)=[\tfrac{3}{8},1+\tfrac{3}{8}[,\quad p=-1;\quad I_1(p+1)=[3.4^{-p-1}/8,3.4^{-p}/8[,\quad p\geqslant 0\quad (m=1), \tag{2.3}$$

$$I_m(0)=[m^{-2}(m+1)^{-2}(2m+1)/2,1-m^{-2}+m^{-2}(m+1)^{-2}(2m+1)/2[\quad (m\geqslant 2), \tag{2.4}$$

$$I_m(p+1)=[m^{-2}(m+1)^{-2}r(m)^{-p-1}(2m+1)/2,m^{-2}(m+1)^{-2}r(m)^{-p}(2m+1)/2[\quad (m\geqslant 2,\ p\geqslant 0), \tag{2.5}$$

the sequence $\{r(m)\}_m$ being such that $r(m) = O(m^3)$ as $m \to \infty$. Then it is possible to construct a sequence $\{g(m)\}_m \downarrow 0$ as $m \to \infty$, with $g(1) = \frac{3}{8}$, such that the interval family $\{I_m(p+1)\}_p$ represents a partition of $J^{(m)} =]0,1+g(m)[$.

(4) Let

$$\mathscr{P}_m(p+1) = \mu[T^{(m)}(\epsilon_0) \cap I_m(p+1)]/\mu[I_m(p+1)] \tag{2.6}$$

be the probability that a randomly chosen $\omega \in]0,1+g(m)[$ belongs to $T^{(m)}(\epsilon_0) \cap I_m(p+1)$, and let $A(\epsilon_0) = 3\epsilon_0^{2/3} - 2\epsilon_0$. Then $\mathscr{P}_m(p+1) \leqslant P_m(p+1)$, where

$$P_1(-1) = A(\epsilon_0)(1+\ln 4); \quad P_1(p+1) = 8A(\epsilon_0)(4+2.4^{p+1}\ln 4)/9, \quad p \geqslant 0, \tag{2.7}$$

$$P_m(-1) = A(\epsilon_0)m^2[1+\ln r(m)](m^2-1)^{-1}, \tag{2.8}$$

$$P_m(p+1) = 2A(\epsilon_0)m^2(m+1)^2[r(m)+2r(m)^{(p+1)}\ln r(m)](2m+1)^{-1}[r(m)-1]^{-1}, \quad p \geqslant 0. \tag{2.9}$$

Remarks: (a) The probability $\mathscr{P}_m(p+1)$ is set to 1 by definition if $P_m(p+1) \geqslant 1$. Furthermore, since $\{I_m(p+1)\}_p$ is a partition, $\mathscr{P}_m(p+1)$ and $P_m(p+1)$ extend to well-defined measures $\omega \to \mathscr{P}_m(\omega)$ and $\omega \to P_m(\omega)$ on $]0,1+g(m)[$.

(b) Since $P_m(p+1)$ is increasing with p, the closer ω is to 0 the larger is the probability of belonging to the "bad" set, i.e., for which the isolation distance is smaller than ϵ_0. It can be shown that the "tail" of $P_m(p+1)$ behaves as ω^{-1} as $\omega \to 0$.

(c) The construction of the pairwise disjoint intervals $I_m(p+1)$ is dictated by the following criterion: let $L_{m,k} \subset]0,1[$ be the smallest interval containing the resonant family $\{\omega(m,n,k)\}_{n \neq m, k \in \mathbb{Z}}$, m fixed. One has

$$L_{1,k} = [\tfrac{3}{4}k, 1/k[;$$

$$L_{m,k} = [(2m+1)m^{-2}(m+1)^{-2}k^{-1}, k^{-1}(1-m^{-2})[, \quad m = 2,\ldots. \tag{2.10}$$

It is easily seen that for any fixed m the intervals $L_{m,k}$ are not pairwise disjoint, while this property might hold for a subsequence $\{L_{m,kp}\}_{p \in \mathbb{N}}$. It will be seen in the proof that the family $\{I_m(p+1)\}$ is essentially a particular subsequence $\{L_{m,kp}\}$, the sequence $\{k_p\}$ being, among all those making the family $\{L_{m,k}\}$ pairwise disjoint, that particular one for which the ratio k_{p+1}/k_p is the smallest integer depending only on m.

(d) It will be also seen that the pairwise disjoint family $\{I_m(p+1)\}$ is the finest possible decomposition. This means that each interval $L_{m,k}$ has a nonempty intersection with at most two intervals in $\{I_m(p+1)\}$. Since the family $\{L_{m,k}\}$ is not pairwise disjoint, this implies that the partition $\{I_m(p+1)\}$ yields the best probabilistic estimates.

Corollary I: (1) For any fixed value β of the probability $\mathscr{P}^{(m)}(\omega)$ there is a step function $\epsilon_0 \to \omega_c(\epsilon_0,\beta,m)$ such that $p^{(m)} \leqslant \beta$ if $\omega \geqslant \omega_c$ and $P^{(m)} > \beta$ if $\omega < \omega_c$.

(2) The step function $\epsilon_0 \to \omega_c$ admits the following monotone continuous majorizations:

$$\omega_c(\epsilon_0,\beta,m) \leqslant 3(16\ln 4)\epsilon_0^{2/3}/8(3\beta - 32\epsilon_0^{2/3}), \quad m = 1, \tag{2.11}$$

$$\omega_c(\epsilon_0,\beta,m) \leqslant \epsilon_0^{2/3}(2m+1)[6\ln r(m)] \times \{[\beta(r(m)-1)(2m+1)] - 6r(m)m^2(m+1)^2\epsilon_0^{2/3}\}^{-1}, \quad m > 1. \tag{2.12}$$

Remark: The estimates (2.11) and (2.12) can be inverted to yield ϵ_0 as a function of ω, namely to yield the probability of the maximum isolation distance for any given ω:

$$\epsilon_c(\beta,\omega,1) = (3\omega\beta)^{3/2}(4\ln 4 + 4\omega)^{3/2}, \quad m = 1, \tag{2.13}$$

$$\epsilon_c(\beta,\omega,m) = \{\beta\omega[r(m)-1] \times (2m+1)\}^{3/2}\{6(2m+1)\ln r(m) + 6r(m)m^2(m+1)^2\omega\}^{-3/2}, \quad m > 1. \tag{2.14}$$

III. PROOF OF THE RESULTS

Proof of Proposition I: Assertion (1) of Proposition I is an elementary computation. To see assertion (2), let us assume $\omega \in B^{(m)} =]0,1[\setminus\omega(m,n,k)$. Then

$$\inf_{n,k}|\lambda_{m,0} - \lambda_{n,k}(\omega)| \leqslant |\lambda_{m,0} - \lambda_{n_1,k_1}(\omega)| = |n^{-2} + k_1\omega - m^{-2}|, \tag{3.1}$$

for some fixed (n_1,k_1). If $\{\omega_j\}$ is a sequence of points of $B^{(m)}$ such that ω_j tends to $\omega(m,n_1,k_1)$ as $j \to \infty$, then $|\lambda_{m,0} - \lambda_{n1,k1}(\omega_j| \to 0$ as $j \to \infty$. Therefore there is no strictly positive lower bound for the isolation distance valid for all $\omega \in B^{(m)}$, and by Lemma 1 in Appendix (2) is proved. To prove assertion (3), given $\epsilon_0 > 0$, it is enough to choose $B^{(m)}(\epsilon_0)$ in the following way:

$$B^{(m)}(\epsilon_0) = \{\omega \in B^{(m)}: \delta(\omega) \equiv \inf_{n,k}|\lambda_{m,0} - \lambda_{n,k}(\omega)| \geqslant \epsilon_0\}. \tag{3.2}$$

This concludes the proof of Proposition I.

Let us now turn to the proof of Proposition II. Consider the open interval $E_{m,n,k}(\epsilon_0)$ of width $2\epsilon_0/k$ centered around each resonant point $\omega(m,n,k)$:

$$E_{m,n,k}(\epsilon_0) =]\omega(m,n,k) - \epsilon_0/k, \omega(m,n,k) + \epsilon_0/k[. \tag{3.3}$$

By definition of $T^{(m)}(\epsilon_0)$ we have

$$\bigcup_{n,k}^{\infty} E_{m,n,k}(\epsilon_0) = T^{(m)}(\epsilon_0) \tag{3.4}$$

and therefore $B^{(m)}(\epsilon_0) = B^{(m)} \setminus T^{(m)}(\epsilon_0)$. We thus see that the problem is to determine the "size" of the union (3.4). To this end it is useful to define the concept of *resonant family of order k*. Given the sequence $\omega(m,n,k) = k^{-1}(n^{-2} - m^{-2})$, $k \in \mathbb{N}$, $n = 2,3,\ldots$ (if $k < 0$, it is enough to consider the sequence $-\omega$), the resonant family of order k is the sequence of semiopen intervals

$$L_{m,k} = [k^{-1}q(m), k^{-1}v(m)[, \quad m>1;$$
$$L_{1,k} = [3/4k, 1/k[, \tag{3.5}$$
$$q(m) = (2m+1)m^{-2}(m+1)^{-2};$$
$$v(m) = 1 - m^{-2}, \quad v(1) = 1. \tag{3.6}$$

It is immediately seen that $L_{m,k}$ is the smallest right-open interval containing, for each fixed k, the whole sequence of resonant points $\{\omega(m,n,k)\}_n$, $n \neq m$. If the resonant families were pairwise disjoint, i.e., $L_{m,k} \cap L_{m,h} = \varnothing$ for $h \neq k$, m fixed, then the estimate of $\mu[T^{(m)}(\epsilon_0)]$ would be trivial. However, it is easy to see that $L_{m,k} \cap L_{m,h} = \varnothing$ if and only if $m=1$ and $(h,k) \in \{1,2,3,4\}$, i.e., the resonant families are *interacting* for $k>4$ ($m=1$) and for all k ($m>1$). We have thus to take into account the intersections, and this makes the problem of estimating the measure of $T^{(m)}(\epsilon_0)$ a not immediately obvious one.

Proof of Proposition II [Assertions (1) and (2)]: (a) To enlighten the main points of the argument, consider first the noninteracting resonant family $L_{1,1} = [\frac{3}{4}, 1.4[$, and the corresponding sequence of resonant points $\omega(1,n,1) = \{1 - n^{-2}\}_{n>1}$. We have $d_1(n) \equiv |\omega(1,n,1) - \omega(1,n+1,1)| = q(n) \sim 2n^{-3}$ as $n \to \infty$. Therefore if $2 \leqslant n \leqslant N(\epsilon_0) \equiv \epsilon_0^{-1/3}$ the distance between two consecutive resonant points exceeds $2\epsilon_0$, and hence for these values of n the intervals $E_{1,n,1}(\epsilon)$ are pairwise disjoint, while this is not true for $n > N(\epsilon_0)$. Therefore if $\omega \in]1 - N^{-2}, 1[$ there is $\omega(1,n_1,1)$ such that $|\omega - \omega(1,n_1,1)| < 2\epsilon_0$ and consequently $|\lambda_{1,0} - \lambda_{n1,1}(\omega)| \leqslant \epsilon_0$, so that $]1 - N^{-2}, 1[$ is contained in $T^{(1)}(\epsilon_0)$. Hence the contribution of $L_{1,1}$ to $T^{(1)}(\epsilon_0)$ is

$$T^{(1)}(\epsilon_0) \cap L_{1,1} = \bigcup_{n=2}^{N(\epsilon_0)} E_{1,n,1}(\epsilon_0) \cup]1 - N^2, 1[\cup]1, 1+\epsilon_0[. \tag{3.7}$$

The interval $[1, 1+\epsilon_0[$ has been included to overestimate the resonant points accumulating at $\omega = 1$. Therefore we have

$$\mu[T^{(1)}(\epsilon_0) \cap L_{1,1}] = 1 - (1 - N^{-2}) + 2\epsilon_0 + (N-2)2\epsilon_0 = 3\epsilon_0^{2/3} - 2\epsilon_0 \equiv A(\epsilon_0). \tag{3.8}$$

Of course we can immediately obtain an upper bound to $\mu[T^{(1)}(\epsilon_0)]$ considering each resonant family $L_{1,k}$ as noninteracting; however this yields $\mu[T^{(1)}(\epsilon_0) \cap L_{1,k}] = k^{-1}A(\epsilon_0)$, and the sum over k is divergent. Hence a better estimate is required for the contribution to $T^{(m)}(\epsilon_0)$ of the infinite number of resonant families $L_{m,k}$.

(b) To this end, first remark that the distance between two consecutive points of the resonant family $L_{m,k}$ is

$$d_k(n) = |\omega(m,n,k) - \omega(m,n+1,k)| = k^{-1}q(n) \sim 2k^{-1}n^{-3} \quad \text{as } n \to \infty. \tag{3.9}$$

Requiring $d_k(n) \leqslant 2k^{-1}\epsilon_0$ we get $n \geqslant N(\epsilon_0) = \epsilon_0^{-1/3}$ for all families $L_{m,k}$. We can now define the interval $\mathscr{A}_{m,k} = [k^{-1}m^2 - N^{-2}, k^{-1}m^{-2}[$ to be the "*accumulation interval*" of the family $L_{m,k}$, because it is formed by all those values of ω such that the isolation distance of $\lambda_{m,0}$ is less than ϵ_0. Remark furthermore that only the intervals $E_{m,n,k}(\epsilon_0)$ centered in the resonant points $\omega(m,n,k)$ for $n = 2,\ldots,N(\epsilon_0) - 1$ are pairwise disjoint. These resonant points will be referred to as *resonant isolated points*, or briefly *isolated points*.

Consider now the accumulation intervals $\mathscr{A}_{m,k}$ and $\mathscr{A}_{m,k+1}$ of two consecutive (*adjacent*) resonant families $L_{m,k}$ and $L_{m,k+1}$. We have

$$\mathscr{A}_{m,k} = [k^{-1}m^{-2} - k^{-1}N^{-2}, k^{-1}m^{-2}[,$$
$$\mathscr{A}_{m,k+1} = [(k+1)^{-1}m^{-2} - (k+1)^{-1}N^{-2}, (k+1)^{-1}m^{-2}[. \tag{3.10}$$

Look at those $\mathscr{A}_{m,k}$ which have nonempty intersection. This happens if $(k+1)^{-1}m^{-2} > k^{-1}m^{-2} - k^{-1}N^{-2}$, i.e.,

$$k \geqslant K_m(\epsilon_0) \equiv m^{-2}(N(\epsilon_0)^2 - m^2). \tag{3.11}$$

By the superposition of the accumulation intervals of the adjacent families the whole interval $[0, K^{-1}[$ is clearly contained in $T^{(m)}(\epsilon_0)$. We have

$$\mu(\,]0, K_m(\epsilon_0)^{-1}[\,) = m^2(\epsilon_0^{-2/3} - m^2)^{-1}. \tag{3.12}$$

To complete the estimate of the measure of the union over $k \geqslant K_m(\epsilon_0)$ of $\mathscr{A}_{m,k}$ we must take into account also all point greater than $K_m^{-1}m^{-2}$. To this end, determine H_m by requiring

$$H_m^{-1}(1 - m^{-2}) = K_m^{-1}m^{-2} \Rightarrow H_m = K_m(m^2 - 1). \tag{3.13}$$

Hence the measure of this last set is not greater than

$$\sum_{k=K_m}^{H_m} \epsilon_0 k^{-1}(m-1) \leqslant 2\epsilon_0(m-1)[K_m^{-1} + \ln(m^2 - 1)]. \tag{3.14}$$

Denoting now by $\mathscr{B}_m$ the union over $k \geqslant K_m(\epsilon_0)$ of $\mathscr{A}_{m,k}$, we have

$$\mu(\mathscr{B}_m) \leqslant m^{-2}K_m^{-1} + 2\epsilon_0(m-1)[K_m^{-1} + \ln(m^2-1)] \tag{3.15}$$

and since the families $\mathscr{A}_{m,k}$ are disjoint for $k \leqslant K_m - 1$,

$$\mu[T^{(m)}(\epsilon_0)] \leqslant \sum_{k=1}^{K_m - 1} A(\epsilon_0)k^{-1} + [m^2K_m]^{-1} + 2\epsilon_0[K_m^{-1} + \ln(m^2-1)] \leqslant F_m(\epsilon_0), \tag{3.16}$$

where, by (3.11) and recalling that $N(\epsilon_0) = \epsilon_0^{-1/3}$:

$$F_m(\epsilon_0) = A(\epsilon_0)[1 + \ln(\epsilon_0^{-2/3} - m^2)] + m^{-2}(\epsilon_0^{-2/3} - m^2)^{-1} + 2\epsilon_0(m-1) \times [m^{-2}\epsilon_0^{-2/3} - 1 + \ln(m^2-1)], \quad m>1, \tag{3.17}$$

$$F_1(\epsilon_0) = A(\epsilon_0)[1 + \ln(\epsilon_0^{-2/3} - 2)] + (\epsilon_0^{-2/3} - 1), \quad m=1. \tag{3.18}$$

Since $F_m(\epsilon_0)$ vanishes as $\epsilon_0 \to 0$, we have $\mu[T^{(m)}(\epsilon_0)] \to 0$ as $\epsilon_0 \to 0$ and therefore $\mu[B^{(m)}(\epsilon_0)] \to \mu(B^{(m)})$ as $\epsilon_0 \to 0$. This proves assertion (1). Assertion (2) is proved if we take $\omega_0(\epsilon_0, m) = K_m(\epsilon_0)^{-1}$.

Remark: A lower bound for $\mu[T^{(m)}(\epsilon_0)]$ is of course

$$G_m(\epsilon_0) = N(\epsilon_0)^{-2}\ln[K_m(\epsilon_0)] = \epsilon_0^{2/3}\ln(m^{-2}\epsilon_0^{-2/3} - 1). \quad (3.19)$$

Note that the ratio $G_m(\epsilon_0)/F_m(\epsilon_0)$ has a constant limit as $\epsilon_0 \to 0$.

Proof of Proposition II [Assertions (3) and (4)]: Consider again the families $L_{m,k}$ given by (3.9) and define the *enlarged resonant families*

$$L_{m,k}(\epsilon_0) = [(q(m) - \epsilon_0)/k,(v(m) + \epsilon_0)/k[, \quad m > 1;$$
$$L_{1,k}(\epsilon_0) = [(\tfrac{3}{4} - \epsilon_0)/k,(1 + \epsilon_0)/k[. \quad (3.20)$$

Let us determine a sequence of positive integers $\{k_p\}_{p>0}$ such that $L_{m,kp}(\epsilon_0) \cap L_{m,kp+1}(\epsilon_0) = \varnothing$. This condition requires $k_{p+1}/k_p \geqslant 4(1+\epsilon_0)/(3-4\epsilon_0)$, $\epsilon_0 < \frac{3}{4}$, for $m = 1$, and $k_{p+1}/k_p \geqslant [v(m)+\epsilon_0]/[q(m)+\epsilon_0]$, $\epsilon_0 \leqslant q(m)$, for $m > 1$. Then we can choose $\{k_p\}$ as a geometric progression of ratio not less than $[4(1+\epsilon_0)/(3-4\epsilon_0)]$ $(m=1)$, $[v(m)+\epsilon_0]/[q(m)+\epsilon_0]$ $(m>1)$. Note that the restriction $\epsilon_0 < q(m)$ is equivalent to the requirement $\epsilon_0 < |\lambda_{m,0} - \lambda_{m+1,0}|$. Taking $\epsilon_0 \leqslant q(m)/2$ one has

$$k_{p+1}/k_p \geqslant s(m), \quad s(m) = 2[v(m) + q(m)/2]/q(m) \quad (3.21)$$

and thus we can take $k_p = r(m)^p$, where

$$r(m) = [s(m)], \quad [x] = \text{integer part of } x. \quad (3.22)$$

Remark that $r(m) = O(m^3)$ as $m \to \infty$. It is now easy to see that in the limiting case $k_{p+1}/k_p = s(m)$ the interval family $\{I_m(p+1)\}_{p \geqslant -1}$, defined as

$$I_m(0) = [q(m)/2, 1 + q(m)/2[;$$
$$I_m(p+1) = [q(m)/(2k_{p+1}), q(m)/(2k_p)[, \quad p \geqslant 0, \quad (3.23)$$

is a partition of $]0,1+q(m)[$. Furthermore, all families $L_{m,k}(\epsilon_0)$ (and thus *a fortiori* all families $L_{m,k}$) have nonempty intersection with at most two elements of the partition. If the ratio is larger than the limiting one the family $\{I_m(p+1)\}_{p \geqslant -1}$ is still a partition of $]0,1+q(m)[$ but it might happen that for some k and p

$$L_{m,k}(\epsilon_0) \subset I_m(p+1). \quad (3.24)$$

This is actually the general situation because the ratio k_{p+1}/k_p is not in general an integer. Then we can distinguish two subcases (A) $L_{m,k}(\epsilon_0) \subset I_m(p+1)$, $\epsilon_0 < q(m)$, for some k and p; and (B) $L_{m,k}(\epsilon_0) \cap I_m(p+1) \neq \varnothing$ for some (k,p), but there are no (p,k) such that $I_m(p+1) \supset L_{m,k}(\epsilon_0)$. Note that $g(m) \to \infty$ as $m \to \infty$, and that $\mu[I_m(p+1)] \to 0$ as $p \to \infty$ for any fixed m. Our problem is now the following: given any interval $I_m(p+1)$ of the partition determine those families $L_{m,k}(\epsilon_0)$ having nonempty intersection with it. In case (A), we must simultaneously have

$$\frac{q(m)}{2k_{p+1}} < \frac{q(m) - \epsilon_0}{k}; \quad \frac{v(m) + \epsilon_0}{k} < \frac{q(m)}{2k_p},$$

i.e., (3.25)

$$\frac{2[v(m)+\epsilon_0]}{q(m)} k_p < k < \frac{2[q(m)-\epsilon_0]}{q(m)} k_{p+1}.$$

In case (B) we must have

$$\frac{2[q(m)-\epsilon_0)]}{q(m)} k_p \leqslant k \leqslant \frac{2[q(m)-\epsilon_0]}{q(m)} k_{p+1},$$

or (3.26)

$$\frac{2[v(m)+\epsilon_0]}{q(m)} k_p \leqslant k \leqslant \frac{2[v(m)+\epsilon_0]}{q(m)} k_{p+1}.$$

Combining this with (3.26) we get that $L_{m,k} \cap I_m(p+1) \neq \varnothing$ if

$$\frac{2[q(m)-\epsilon_0]}{q(m)} k_p \leqslant k \leqslant \frac{2[v(m)+\epsilon_0]}{q(m)} k_{p+1}. \quad (3.27)$$

Consider now the probability that $\omega \in I_m(p+1)$ belongs to $T^{(m)}(\epsilon_0)$:

$$\mathscr{P}_m(p+1) = \mu[I_m(p+1) \cap T^{(m)}(\epsilon_0)]/\mu[I_m(p+1)]. \quad (3.28)$$

To get an upper bound on $\mathscr{P}_m(p+1)$ consider first $p \geqslant 0$ so that

$$\mu[I_m(p+1)] = q(m)[r(m) - 1]/2r(m)^{p+1}. \quad (3.29)$$

Now the contribution of any single $L_{m,k}(\epsilon_0)$ to $\mu[T^{(m)}(e_0)]$ is majorized by $A(\epsilon_0)/k$. Hence by (3.27),

$$\mu[(I_m(p+1)] \cap T^{(m)}(\epsilon_0)] \leqslant A(\epsilon_0) \sum_{2[q(m)-\epsilon_0]k_p/q(m)}^{2[v(m)+\epsilon_0]k_{p+1}/q(m)} k^{-1} \leqslant A(\epsilon_0) \sum_{k=k_p}^{r(m)k_{p+1}} k^{-1}$$
$$\leqslant A(\epsilon_0)[1/k_p + 2\ln r(m)]. \quad (3.30)$$

Therefore

$$\mathscr{P}_m(p+1) \leqslant P_m(p+1) = A(\epsilon_0)[r(m)^{-p} + 2\ln r(m)]r(m)^{p+1}m^2(m+1)^2 \times \{[(r(m)-1)(2m+1)]\}^{-1}. \quad (3.31)$$

If $p = -1$ the families $L_{m,k}$ have nonempty intersection with $I_m(0)$ if $1 \leqslant k \leqslant r(m)$, and $\mu[I_m(0)] = 1 - m^{-2}$. Hence

$$\mathscr{P}_m(0) \leqslant P_m(0) = A(\epsilon_0)[1 + \ln r(m)]m^2(m^2-1)^{-1}, \quad m > 1,$$
$$P_1(p+1) \leqslant \tfrac{8}{9}A(\epsilon_0)[4 + 2.4^{p+1}\ln 4], \quad p \geqslant 0; \quad (3.32)$$
$$P_1(-1) \leqslant A(\epsilon_0)(1 + \ln 4), \quad p = -1 \quad (m=1).$$

This concludes the proof of Proposition II.

Let us now turn to the proof of Corollary I. For $\omega \in I_m(p+1)$, we define $\mathscr{P}_m(\omega) = \mathscr{P}_m(p+1)$. Next remark that $P_m(p+1)$ is monotonically increasing in p; if the estimates have to make sense from a probabilistic point of view, this function cannot exceed 1, which is not true for all values of p (ϵ_0 fixed). We assume therefore $\mathscr{P}_m(\omega) = 1$ if $P_m(p+1) \geqslant 1$. This means that the above estimates cannot exclude the occurrence of an interval $[0,\Omega_0(\epsilon_0,m)[\supset [0,\omega_0(\epsilon_0,m)[$ for which $\mathscr{P}_m(\omega) = 1$. Therefore a more sophisticated analysis is required, aimed at determining those values of ϵ_0 for which $P_m(p+1)$ does not exceed a given quantity.

Proof of Corrollary I: Consider first $P_m(0)$, and require $P_m(0) < 1$. By (3.32) we get $\epsilon_0 < \epsilon_0(m,-1)$, with

$$\epsilon_0(m,-1) = \{3[1+\ln r(m)]m^{-2}(m+1)^{-2}\}^{3/2}, \quad m>1; \tag{3.33}$$

$$\epsilon_0(1,-1) = [3(4\ln 4)]^{3/2}.$$

Hence we first have to take $\epsilon_0<\epsilon_0(m,-1)$ to avoid $\mathscr{P}_m(\omega)=1$ on $[0,1+g(m)[$. With such an ϵ_0 consider $P_m(p+1)$ for $p\geqslant 0$ and require $P_m(p+1)<1$. By (3.31) we have $\epsilon_0<\epsilon_0(m,p)$, with

$$\epsilon_0(m,p) = \{(2m+1)(r(m)-1)[6[r(m) + 2r(m)^{p+1}\ln r(m)]m^2(m+1)^2]^{-1}\}^{3/2}, \quad m>1, \tag{3.34}$$

$$\epsilon_0(1,p) = [4+2.4^{p+1}\ln 4]^{-3/4}. \tag{3.35}$$

We can now determine $\Omega_0(\epsilon_0,m)$. By definition, if $\epsilon=\epsilon_0(m,p)$, $P_m(p+1)=1$, $P_m(p)<1$, $P_m(p+2)>1$. Therefore $\Omega_0(\epsilon_0,m)$ is the right end point of $I_m(p+1)$. If ϵ_0 decreases then $\Omega_0(\epsilon_0,m)=\Omega_0(\epsilon_0(m,p),m)$ for all $\epsilon_0>\epsilon_0(m,p+1)$. Therefore $\epsilon_0\to\Omega_0(\epsilon_0,m)$ is the step function defined in the following way:

$$\Omega_0(\epsilon_0,m) = q(m)r(m)^{-p}/2, \quad \epsilon_0(m,p+1)<\epsilon_0\leqslant\epsilon_0(m,p). \tag{3.36}$$

By the same argument we now prove assertion (1). Fix $b, 0<\beta<1$, and require $P_m(p+1)<\beta$. For $p=-1$ we immediately find the condition

$$\epsilon_0\leqslant\beta^{3/2}[3(1+\ln r(m))m^2(m^2-1)^{-1}]^{-3/2}\equiv\epsilon_0(m,-1,\beta). \tag{3.37}$$

Then if $\epsilon_0>\epsilon_0(m,-1,\beta)$ we have $P_m(0)>\beta$, and thus $\mathscr{P}_m(\omega)<\beta$ if ω is greater than the right end point of $I_m(0)$. Hence $\omega_c(\beta,\epsilon_0,m)=1+q(m)$ if $\epsilon_0>\epsilon_0(m,-1,\beta)$ and this proves (1). For $p\geqslant 0$ we require now $P_m(p+1)<\beta$. By (3.31) this implies the condition

$$\epsilon_0\leqslant\epsilon_0(m,p,\beta)\equiv\beta^{3/2}[(2m+1)(r(m)-1)]^{3/2} \times\{6[r(m)+2r(m)^{p+1}\ln r(m)]m^2(m+1)^2\}^{-3/2}. \tag{3.38}$$

We can again state that if $\epsilon_0>\epsilon_0(m,p+1,\beta)$ then $P_m(p+1)>\beta$, while if $\epsilon_0(m,p+1,\beta)<\epsilon_0\leqslant\epsilon_0(m,p,\beta)$ then $P_m(p+1)<\beta$ and $P_m(p+2)>\beta$. Therefore if ϵ_0 is so small that there is $p\geqslant 0$ such that $\epsilon_0(m,p,+1,\beta)\leqslant\epsilon_0\leqslant\epsilon_0(m,p,\beta)$ then $P_m(p+1)\leqslant\beta$ and $P_m(p+2)\geqslant\beta$. We can thus conclude that ω_c is the left end point of $I_m(p+1)$. Hence the step function $\epsilon_0\to\omega_c(\epsilon_0,\beta,m)$ is given by

$$\omega_c = q(m)r(m)^{p+1}/2, \quad \epsilon_0(m,p+1,\beta)<\epsilon_0<\epsilon_0(m,p+2,\beta). \tag{3.39}$$

Finally to prove the monotone continuous majorization (useful to visualize the behavior in the various parameters) solve for $r(m)^{p+1}$ from definition (3.38) of $\epsilon_0(m,p,\beta)$:

$$r(m)^{p+1} = \frac{\beta[(2m+1)(r(m)-1)] - 6r(m)m^2(m+1)^2\epsilon_0(m,p,\beta)^{2/3}}{12\epsilon_0(m,p,\beta)r(m)^{p+1}m^2(m+1)^2\ln r(m)}. \tag{3.40}$$

Therefore

$$\omega_c(\epsilon_0(m,p,\beta),\beta,m) = \frac{6(2m+1)\ln r(m)\epsilon_0(m,p,\beta)^{2/3}}{\beta[(2m+1)(r(m)-1)] - 6r(m)m^2(m+1)^2\epsilon_0(m,p,\beta)^{2/3}}. \tag{3.41}$$

If we replace the sequence $\{\epsilon_0(m,p,\beta)\}_p$ by the continuous variable ϵ_0 we get

$$\epsilon_0\to\omega_c(\beta,\epsilon_0,m) = \frac{6(2m+1)\ln r(m)\epsilon_0^{2/3}}{\beta[(2m+1)(r(m)-1)] - 6r(m)m^2(m+1)^2\epsilon_0^{2/3}}. \tag{3.42}$$

This proves (2). To see the remark, which yields the maximum value of ϵ_0 statistically allowed for a given ω, it is enough to invert (3.42) with respect to ω_c. Solving for ω_c we get

$$\omega\to\epsilon_0(\omega,\beta,m) = \frac{\omega^{3/2}\beta^{3/2}[(r(m)-1)(2m+1)]^{3/2}}{[6(r(m)-1)m^2(m+1)^2\omega + 6(2m+1)\ln r(m)]^{3/2}}. \tag{3.43}$$

This concludes the proof of Corollary I.

ACKNOWLEDGMENT

One of the authors (S. G.) was partially supported by NSF under Grant No. PHY-8502383. The work of one of the others (E. C.) was partially supported by Ministero Pubblica Istruzione.

APPENDIX: SOME PERTURBATION ESTIMATES

Let us work out in some detail a statement (trivial in the self-adjoint case) essentially contained in Ref. 10, VII § 1 and § 2.

Lemma I: Let $\lambda(\omega)$ be an isolated simple eigenvalue of $K(\theta)$, $0<\text{Im }\theta<\pi/4$, and let $\lambda(\omega,F)$ be the corresponding resonance, i.e., the nearby eigenvalue of $K(F,\theta)$ existing for F suitably small. Then the Rayleigh–Schrödinger expansion of initial point $\lambda(\omega)$ has a positive radius of convergence $r_0(\lambda)$ if and only if the isolation distance $d=d(\lambda)$ is positive.

Proof: Since for $0<\text{Im }\theta<\pi/4$, $K(F,\theta)$ is a type A-holomorphic family with respect to F[8], by Refs. 10, VII § 2.3 we have

$$r_0\geqslant\min_{z\in\Gamma}[a\|R(z,\theta)\| + b\,\|K(\theta)R(z,\theta)\|]^{-1}$$

$$\equiv\{\max_{z\in\Gamma}[a\|R(z,\theta)\| + b\,\|K(\theta)R(z,\theta)\|]\}^{-1}, \tag{A1}$$

where $R(z,\theta)=[K(\theta)-z]^{-1}$ is the free resolvent, Γ is any closed regular complex curve entirely contained in the resolvent set $\rho(K(\theta))$ of $K(\theta)$ and separating $\lambda(\omega)$ from any other point of $\sigma(K(\theta))$, and (a,b) are the relative boundedness constants of W as estimated in Ref. 8. Since all eigenvalues of $K(\theta)$ are semisimple, by Ref. 10, III § 6.5 we have

$$R(z,\theta) = (z-\lambda)^{-1}P_\lambda(\theta) + R'(z,\theta), \qquad (A2)$$

where $P_\lambda(\theta)$ is the projection on the eigenvector corresponding to λ and $R'(z,\theta)$, the reduced resolvent, is holomorphic at $z=\lambda$. Setting $r=|\lambda - z|$ (A2) yields, for $z\in\Gamma = \{z\colon |\lambda - z| = r\}$,

$$r^{-1}\|P_\lambda\| - \|R'(z,\theta)\| \leqslant \|R(z,\theta)\| \leqslant r^{-1}\|P_\lambda\| + \|R'(z,\theta)\|. \qquad (A3)$$

It follows by general arguments on holomorphic functions that $Q(d) = \|R'(\lambda,\theta)\| \to \infty$ only if $d\to 0$. A simple application of the second resolvent formula now yields, for $r<Q(d)$, d fixed,

$$\|R'(z,\theta)_{z\in\Gamma}\| \leqslant \|R'(\lambda,\theta)[1 - rR'(\lambda,\theta)]^{-1}\| \leqslant Q(d)[1+Q(d)]^{-1}. \qquad (A4)$$

If we choose $r=r(d)=[3Q(d)]^{-1}$, i.e., $\Gamma = \{z\in\mathbb{C}\colon |\lambda - z| = r(d)\}$, (A4) yields

$$\|R'(z,\theta)_{z\in\Gamma}\| \leqslant 3Q(d)/2. \qquad (A5)$$

Since P_λ is a one-dimensional projection, we can always assume $\|P_\lambda\| = 1$. By the second of (A3) and (A5) we then have $\|R(z,\theta)_{z\in\Gamma}\| \leqslant 3Q(d) + 3Q(d)/2 = 9Q(d)/2$, and since $K(\theta)R(z,\theta) = \mathbb{1} + zR(z,\theta)$, $|z| = |\lambda - \lambda + z| \leqslant r(d) + |\lambda|$, we can write

$$\|K(\theta)R(z,\theta)_{z\in\Gamma}\| \leqslant 1 + 3[1 + 3|\lambda|\,Q(d)]/2, \qquad (A6)$$

whence, by (A1),

$$r_0 \geqslant 2\{9aQ(d) + b\,[5 + 9|\lambda|\,Q(d)]\}^{-1}.$$

Conversely, let there exist $\omega_0 \in]0,1[$ such that $d(\lambda,\omega)\downarrow 0$ as $\omega\to\omega_0$. It is well known that under these circumstances the limit as $\omega\to\omega_0$ of the Rayleigh–Schrödinger perturbation series does not exist. This proves Lemma I.

[1]See, e.g., the contributions of A. Tip and A. Maquet, in "Proceedings of the Bielefeld encounters in physics and mathematics, VI," *Lecture Notes in Physics*, Vol. 211 (Springer, Berlin, 1984), and the contribution of J. Bellissard, in "Schrödinger operators. Proceedings of the 3rd 1984 Cime session," *Lecture Notes in Mathematics*, Vol. 1159 (Springer, Berlin, 1985).

[2]See, e.g., *Chaotic Behaviour in Quantum Mechanics*, edited by G. Casati (Plenum, New York, 1984).

[3]See, e.g., G. Casati, B. V. Chirikov, D. S. Shepelyanski, and I. Guarneri, to appear in Phys. Rep.

[4]K. Yajima, J. Math. Soc. **29**, 729 (1977).

[5]K. Yajima, Commun. Math. Phys. **87**, 331 (1982).

[6]B. Simon, Ann. Math. **97**, 247 (1973).

[7]M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV* (Academic, New York, 1978).

[8]S. Graffi, V. Grecchi, and H. J. Silverstone, Ann. Inst. H. Poincaré **42**, 215 (1985).

[9]S. Graffi and V. Grecchi, Commun. Math. Phys. **62**, 83 (1978).

[10]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966).

[11]R. V. Jensen, Phys. Rev. A **30**, 386 (1984).

Canonical transformations and exact invariants for dissipative systems

I. A. Pedrosa
Universidade Federal da Paraíba, Departamento de Física-CCEN, 58.000-João Pessoa (PB), Brazil

(Received 4 December 1986; accepted for publication 3 June 1987).

A simple treatment to the problem of finding exact invariants and related auxiliary equations for time-dependent oscillators with friction is presented. The treatment is based on the use of a time-dependent canonical transformation and an auxiliary transformation.

I. INTRODUCTION

In recent years the study of the problem of finding exact invariants for the time-dependent harmonic oscillator (the so-called Ermakov problem) has attracted considerable interest in the literature, both in classical[1–12] and quantum[13–18] mechanics. Apart from its intrinsic mathematical interest, the invariants have invoked much attention because of their use in discussing several physical problems. For example, exact invariants have been applied to the description of the motion of a charged particle in a time-dependent electromagnetic field,[13] to construct coherent states for certain time-dependent systems with several possible physical applications,[19] and to construct exact time-dependent solutions of the Vlasov–Poisson equations.[20]

Essentially, the Ermakov problem consists of demonstrating that, for the time-dependent Hamiltonian,

$$H = p^2/2m + [m\omega^2(t)/2]q^2, \tag{1.1}$$

where q and p are canonically conjugate and $\omega(t)$ is the time-dependent harmonic oscillator frequency, there exists a time-dependent invariant (the Ermakov–Lewis invariant)

$$I = \tfrac{1}{2}[(q\dot{\alpha} - \alpha\dot{q})^2 + (q/\alpha)^2], \tag{1.2}$$

where $q(t)$ satisfies the harmonic oscillator equation

$$\ddot{q} + \omega^2(t)q = 0, \tag{1.3}$$

which is obtained from (1.1) via Hamilton's equations and $\alpha(t)$ is any solution of the auxiliary equation

$$\ddot{\alpha} + \omega^2(t)\alpha = 1/\alpha^3, \tag{1.4}$$

where overdots indicate differentiation with respect to time. The pair of equations (1.3) and (1.4) is called an Ermakov system.

The solution of the Ermakov problem and their generalizations have been found by various methods. The usual methods are (1) Ermakov's method,[3,21] (2) Kruskal's method of exact adiabatic invariants,[13,22] (3) Leach's method of time-dependent canonical transformations,[23,24] (4) Noether's theorem,[2,25] (5) the Lie theory of extended groups,[26,27] and (6) Nelson's stochastic mechanics.[28] These methods have also been applied in search of invariants for dissipative systems (Ermakov systems with friction) using different approaches.[2,12,17,28,29]

The main purpose of the present paper is to exhibit, in a very simple way, an alternative treatment to the problem of finding exact invariants and related auxiliary equations for dissipative systems. The treatment is based on the use of a time-dependent canonical transformation, which reduces the time-dependent oscillator with friction to one without friction with a modified frequency, and an auxiliary time-dependent transformation. Then, employing these two transformations, the Ermakov system with friction and its corresponding invariant are derived directly from the transformed Ermakov problem about which there exists extensive literature.

A brief outline of the present paper is as follows. In Sec. II, we outline our treatment by considering an appropriated time-dependent Hamiltonian. In Sec. II, we apply the treatment for a more general time-dependent Hamiltonian. Finally, some concluding remarks are added in Sec. IV.

II. THE TREATMENT

We start with the time-dependent harmonic-oscillator Hamiltonian

$$H = f(t)(p^2/2m) + f^{-1}(t)\,[m\omega^2(t)/2]q^2, \tag{2.1}$$

where $f(t)$ is an arbitrary real function of time t. The Hamilton equations are

$$\dot{q} = f(t)p/m, \tag{2.2}$$

$$\dot{p} = -f^{-1}(t)m\omega^2(t)q. \tag{2.3}$$

The equation of motion obtained is

$$\ddot{q} + \gamma(t)\dot{q} + \omega^2(t)q = 0, \tag{2.4}$$

where

$$\gamma(t) = -\frac{d}{dt}[\ln f(t)] \tag{2.5}$$

is the time-dependent coefficient friction. Note that the well-known Kanai–Caldirola Hamiltonian[30] is recovered when $f(t) = \exp(-\gamma t)$ with constant γ.

To find an Ermakov–Lewis-type invariant for the Hamiltonian (2.1) we proceed as follows. Consider the time-dependent canonical transformation given by the generating function

$$F(q,P,t) = qPf^{-1/2}(t) - [m\gamma(t)/4]q^2 f^{-1}(t). \tag{2.6}$$

The transformation equations are $Q = \partial F/\partial P$, $p = \partial F/\partial q$, from which we obtain the new canonical variables

$$Q = qf^{-1/2}(t), \tag{2.7}$$

$$P = pf^{1/2}(t) + [m\gamma(t)/2]qf^{-1/2}(t). \tag{2.8}$$

This is a generalization of the canonical transformation proposed by Gzyl.[31] Then, under this transformation the Hamiltonian (2.1) is transformed into a new Hamiltonian $H_1 = H + \partial F/\partial t$ which, in terms of the new variables, is expressed as

$$H_1 = P^2/2m + [m\Omega^2(t)/2]Q^2, \tag{2.9}$$

where

 0022-2488/87/112662-03$02.50

$$\Omega^2(t) = \omega^2(t) - (\gamma^2(t)/4 + \dot{\gamma}(t)/2) \qquad (2.10)$$

is the modified frequency. Here we observe that the Hamiltonian (2.9) is of the form (1.1). Hence an exact invariant for (2.9) is given by

$$I = \tfrac{1}{2}\,[(Q\dot{\rho} - \rho\dot{Q})^2 + (Q/\rho)^2]\,, \qquad (2.11)$$

where $Q(t)$ satisfies the equation of motion

$$\ddot{Q} + \Omega^2(t)Q = 0, \qquad (2.12)$$

and $\rho(t)$ satisfies the auxiliary equation

$$\ddot{\rho} + \Omega^2(t)\rho = 1/\rho^3. \qquad (2.13)$$

We now introduce the transformation

$$\rho(t) = \alpha(t) f^{-1/2}(t), \qquad (2.14)$$

where $\alpha(t)$ is an function of time t to be determined. Then, using (2.7), (2.10), and (2.14) the equation of motion (2.12) is converted into Eq. (2.4) and the auxiliary equation (2.13) into the equation

$$\ddot{\alpha} + \gamma(t)\dot{\alpha} + \omega^2(t)\alpha = f^2(t)/\alpha^3. \qquad (2.15)$$

The Ermakov–Lewis invariant (2.11) is converted into the form

$$I = \tfrac{1}{2}\{f^{-2}(t)\,[(q\dot{\alpha} - \alpha\dot{q})^2] + (q/\alpha)^2\}\,. \qquad (2.16)$$

Thus the pair of equations (2.4) and (2.15) constitute an Ermakov system for the Hamiltonian (2.1) with an invariant of the form (2.16). For $f(t) = 1$ we recover the invariant (1.2). Note that in this case the function (2.6) generates the identity transformation. We also see that an invariant for the Kanai–Caldirola Hamiltonian is just a special case of (2.16) with $f(t) = \exp(-\gamma t)$ with constant γ.

At this point, we remark that in the Ermakov problem one must know both the equation of motion and the auxiliary equation before one can derive the Ermakov–Lewis invariant. But for a general equation of motion there may be an infinite number of different auxiliary equations. Hence a given time-dependent Hamiltonian can possess many different invariants. For example, if we consider instead of Eq. (2.13) the auxiliary equation

$$\ddot{\rho} + \Omega^2(t)\rho = 0, \qquad (2.17)$$

the invariant for (2.9) takes the form[25]

$$I = \tfrac{1}{2}\,[(Q\dot{\rho} - \rho\dot{Q})^2]\,. \qquad (2.18)$$

Now by using (2.7), (2.10), and (2.14) we convert (2.17) into the equation

$$\ddot{\alpha} + \gamma(t)\dot{\alpha} + \omega^2(t)\alpha = 0 \qquad (2.19)$$

and the invariant (2.18) to the form

$$I = \tfrac{1}{2}\{f^{-2}(t)\,[(q\dot{\alpha} - \alpha\dot{q})^2]\}. \qquad (2.20)$$

So, (2.4) and (2.19) constitute other Ermakov systems for (2.1) with an invariant given by (2.20). Note that the pair of equations (2.4) and (2.19) are equivalent to two uncoupled time-dependent harmonic oscillators with friction.

As a further example consider the auxiliary equation[3,5]

$$\ddot{\rho} + \Omega^2(t)\rho = (1/Q\rho^2)\,W(Q/\rho), \qquad (2.21)$$

where $Q(t)$ is a solution of Eq. (2.12) and $W(Q/\rho)$ is an arbitrary function of Q/ρ. For this case the invariant is expressed as[3,5]

$$I = \frac{1}{2}\left[(Q\dot{\rho} - \rho\dot{Q})^2 + 2\int^{Q/\rho} W(u)\,du\right]. \qquad (2.22)$$

We again use (2.7), (2.10), and (2.14) to transform Eq. (2.21) into the equation

$$\ddot{\alpha} + \gamma(t)\dot{\alpha} + \omega^2(t)\alpha = (f^2(t)/q\alpha^2)W(q/\alpha) \qquad (2.23)$$

and the invariant (2.22) to the form

$$I = \frac{1}{2}\left\{f^{-2}(t)\,[(q\dot{\alpha} - \alpha\dot{q})^2] + 2\int^{q/\alpha} W(u)\,du\right\}. \qquad (2.24)$$

We thus see that the pair of equations (2.4) and (2.23) represent another Ermakov system described by the Hamiltonian (2.1). The corresponding invariant is given by expression (2.24). Notice that for $W = q/\alpha$ expression (2.24) becomes the invariant (2.16). We also note that for $W = 0$ the invariant (2.24) reduces to the invariant (2.20).

III. APPLICATION TO NONHARMONIC SYSTEMS

We now consider the time-dependent nonharmonic Hamiltonian

$$H = f(t)\,\frac{p^2}{2m} + f^{-1}(t)\,\frac{m\omega^2(t)}{2}\,q^2 + \frac{mf(t)}{\alpha^2}\,g\!\left(\frac{\alpha}{q}\right), \qquad (3.1)$$

where $\alpha(t)$ is an function of time t to be determined later and $g(\alpha/q)$ is an arbitrary function of α/q. The equation of motion for q follows from (3.1) and is expressed as

$$\ddot{q} + \gamma(t)\dot{q} + \omega^2(t)q = [f^2(t)/\alpha q^2]G(\alpha/q), \qquad (3.2)$$

where $\dot{q}$ and $\gamma(t)$ are given, respectively, by (2.2) and (2.5) and $G = dg/d(\alpha/q)$.

To obtain an exact invariant for (3.1), we proceed as in Sec. II. We transform the Hamiltonian (3.1) to the new Hamiltonian

$$H_2 = \frac{P^2}{2m} + \frac{m\Omega^2(t)}{2}\,Q^2 + \frac{m}{\rho^2}\,g\!\left(\frac{\rho}{Q}\right), \qquad (3.3)$$

where $\Omega(t)$ is given by (2.10). Note that to arrive at the form (3.3) we have also employed the auxiliary transformation (2.14). Now, it is well known that an invariant for (3.5) has the form [3–5]

$$I = \tfrac{1}{2}\,[(Q\dot{\rho} - \rho\dot{Q})^2 + (Q/\rho)^2 + 2g(\rho/Q)]\,, \qquad (3.4)$$

where $Q(t)$ and $\rho(t)$ satisfy, respectively,

$$\ddot{Q} + \Omega^2(t)Q = (1/\rho Q^2)G(\rho/Q) \qquad (3.5)$$

and

$$\ddot{\rho} + \Omega^2(t)\rho = 1/\rho^3. \qquad (3.6)$$

Now, following the same steps as those of Sec. II, we convert the invariant (3.4) into the form

$$I = \tfrac{1}{2}\{f^{-2}(t)[(q\dot{\alpha} - \alpha\dot{q})^2] + (q/\alpha)^2 + 2g(\alpha/q)\}\,. \qquad (3.7)$$

Equations (3.5) and (3.6) are converted into the equations

$$\ddot{q} + \gamma(t)\dot{q} + \omega^2(t)q = [f^2(t)/\alpha q^2]G(\alpha/q) \qquad (3.8)$$

and

$$\ddot{\alpha} + \gamma(t)\dot{\alpha} + \omega^2(t)\alpha = f^2(t)/\alpha^3, \qquad (3.9)$$

which are precisely the pair of equations (3.2) and (2.15). We recognize (3.8) and (3.9) as an Ermakov system described by (3.1) with the Ermakov–Lewis invariant (3.7). Note that for $g = 0$ the invariant (3.7) reduces to the invariant (2.16).

A simple generalization of this result is obtained if we consider instead of Eq. (3.6) the auxiliary equation (2.21),

$$\ddot{\rho} + \Omega^2(t)\rho = (1/Q\rho^2)W(Q/\rho). \qquad (3.10)$$

For this case, the invariant is given by[3–5]

$$I = \frac{1}{2}\left[(Q\dot{\rho} - \rho\dot{Q})^2 + 2\int^{Q/\rho} W(u)du + 2g\left(\frac{\rho}{Q}\right)\right], \qquad (3.11)$$

which can be transformed to the form

$$I = \frac{1}{2}\left\{f^{-2}(t)[(q\dot{\alpha} - \alpha\dot{q})^2] + 2\int^{q/\alpha} W(u)du + 2g\left(\frac{\alpha}{q}\right)\right\}, \qquad (3.12)$$

where $q(t)$ satisfies Eq. (3.8) and $\alpha(t)$ satisfies Eq. (2.23),

$$\ddot{\alpha} + \gamma(t)\dot{\alpha} + \omega^2(t)\alpha = (f^2(t)/q\alpha^2)W(q/\alpha). \qquad (3.13)$$

Note that for $W = q/\alpha$ the invariant (3.12) becomes the invariant (3.7). We also notice that the results (3.7) and (3.12) represent a generalization of the so-called Ray–Reid invariants that have been obtained in other ways by some authors.[4,5,8,15]

IV. CONCLUDING REMARKS

In this paper we have outlined an alternative treatment to find exact invariants and related auxiliary equations for harmonic and nonharmonic dissipative systems. We have seen that this treatment provides, contrary to those employed by some authors,[4,5,8,15] a direct and unsophisticated generalization of the Ray–Reid invariants. Furthermore, it allows a straightforward transition from classical to quantum physics since the transformation determined by (2.6) corresponds to a unitary transformation.[31] Also, it would be interesting to compare the above treatment with those developed by Leach[23,29] and Reid and Ray.[5]

In conclusion, we mention that quantum solutions of Ermakov systems described by the Hamiltonians (1.1) and (3.1) with $f(t) = 1$ are well known in the literature.[32,33] Then, in principle, our treatment could be applied to obtain quantum solutions for Ermakov systems described by the Hamiltonians (2.1) and (3.1). Here we note that quantum solutions for Ermakov systems with friction have been rarely explored in the literature, to the best of the author's knowledge. We also observe that Hartley and Ray[19] have used the invariant (1.2) to construct coherent states for the time-dependent Hamiltonian (1.1). Thus it seems that there would not be any problems constructing coherent states for the Hamiltonian (2.1) using the same technique as presented in Ref. 19. We hope to report on these two possibilities in the future.

Note added in proof: In the discussion about different auxiliary equations (see Sec. II), we have mentioned that a given time-dependent Hamiltonian can possess many different invariants. However, for a one-dimensional problem they will be related via some transformation. In other words, the many different forms of the invariant are equivalent. By way of example, if one considers the auxiliary equations (2.13) and (2.17) and takes $\rho(t)$ to be the solution of (2.13), then the solution of (2.17) is

$$A\rho \sin T + B\rho \cos T,$$

where A and B are constants and

$$T = \int^t \rho^{-2}(\eta)d\eta.$$

ACKNOWLEDGMENTS

The author wishes to thank the referee for supplying the contents of the note added.

[1] C. J. Eliezer and A. Gray, SIAM J. Appl. Math **30**, 463 (1976).
[2] M. Lutzky, Phys. Lett. A **68**, 3 (1978).
[3] J. R. Ray and J. L. Reid, Phys. Lett. A **71**, 317 (1979).
[4] M. Lutzky, J. Math. Phys. **21**, 1370 (1980).
[5] J. L. Reid and J. R. Ray, J. Math. Phys. **21**, 1583 (1980).
[6] J. R. Ray, J. L. Reid, and M. Lutzky, Phys. Lett. A **84**, 42 (1981).
[7] W. Sarlet, Phys. Lett. A **82**, 161 (1981).
[8] J. L. Reid and J. R. Ray, J. Math. Phys. **23**, 503 (1982).
[9] R. K. Colegrave and M. S. Abdalla, J. Phys. A **16**, 3805 (1983).
[10] P. G. L. Leach, Phys. Lett. A **98**, 89 (1983).
[11] I. C. Moreira, J. Phys. A **18**, 899 (1985).
[12] M. R. Feix and H. R. Lewis, J. Math. Phys. **26**, 68 (1985).
[13] H. R. Lewis, Jr. and W. B. Reisenfeld, J. Math. Phys. **10**, 1458 (1969).
[14] B. Remaud and E. S. Hernandez, Physica (Utrecht) A **103**, 35 (1980).
[15] L. S. Wollenberg, Phys. Lett. A **79**, 269 (1980).
[16] R. Vaidyanathan, Phys. Lett. A **84**, 93 (1981); J. Math. Phys. **23**, 1346 (1982).
[17] R. S. Kaushal and H. J. Korsch, J. Math. Phys. **22**, 1904 (1981).
[18] A. K. Dhara and S. V. Lawande, J. Math. Phys. **27**, 1331 (1986).
[19] J. G. Hartley and J. R. Ray, Phys. Rev. D **25**, 382 (1982); J. R. Ray, *ibid.*, **25**, 3417 (1982).
[20] H. R. Lewis and K.R. Symon, Phys. Fluids **27**, 192 (1984).
[21] V. P. Ermakov, Univ. Izv. Kiev **20**(9), 1 (1980).
[22] W. Sarlet, J. Phys. A **11**, 843 (1978).
[23] P. G. L. Leach, J. Math. Phys. **18**, 1608 (1977); **20**, 96 (1979); J. Phys. A **16**, 3261 (1983).
[24] H. R. Lewis and P. G. L. Leach, J. Math. Phys. **23**, 165, 2371 (1982).
[25] J. R. Ray and J. L. Reid, J. Math. Phys. **22**, 91 (1981).
[26] P. G. Leach, J. Math. Phys. **22**, 465 (1981).
[27] J. L. Reid and J. R. Ray, J. Phys. A **15**, 2751 (1982).
[28] A. B. Nassar, Phys. Rev. A **32**, 1862 (1985); J. Math. Phys. **27**, 755 (1986).
[29] P. G. L. Leach, SIAM J. Appl. Math. **34**, 496 (1978).
[30] E. Kanai, Prog. Theor. Phys. **3**, 440 (1948); P. Caldirola, Nuovo Cimento **18**, 393 (1941); see also H. Dekker, Phys. Rep. **80**, 1 (1981).
[31] H. Gzyl, Phys. Rev. A **27**, 2297 (1983).
[32] J. G. Hartley and J. R. Ray, Phys. Rev. A **24**, 2873 (1981); **25**, 2388 (1982).
[33] J. R. Ray, Phys. Rev. D **25**, 3417 (1982); Phys. Rev. A **26**, 729 (1982).

Evaporation of nonzero rest mass particles from a black hole

J. Diaz Bejarano and C. Miro Rodriguez
Departamento de Física, Universidad de Extremadura, 06071-Badajoz, Spain

(Received 7 August 1986; accepted for publication 24 June 1987)

Analytic expressions for the transmission coefficient and the emission and the absorption rates for scalar particles with mass and a chargeless, nonrotating black hole are calculated by using Jacobian elliptic functions and integrals in the Jeffreys–Wentzel–Kramers–Brillouin (JWKB) approximation.

I. INTRODUCTION

In a paper fundamental for the relation of quantum mechanics to general relativity, Hawking[1] showed that a black hole will emit particles like a black body with temperature T inversely proportional to mass M. Many authors, using different arguments, have confirmed this prediction for the Schwarzschild, Kerr, Kerr–Newman, and Vaidya,[2–5] metrics. For an uncharged, nonrotating black hole, the expected number of particles of a given kind, with energy ω is, in Planck units,[1]

$$N = \Gamma[\exp(8\pi M\omega) \pm 1]^{-1}. \tag{1}$$

Here the minus sign is for bosons and the plus sign is for fermions, and Γ is the transmission coefficient of the hole for a particle of rest mass μ.

The transmission coefficient is important for determining black-hole mechanics and thermodynamics: total emitted power and torque, rate of entropy production, etc. However, the calculation is difficult. Most authors take the approximation $\Gamma = 0$ or $\Gamma = 1$ when the particle has an energy either above or below the potential barrier. Thus Dewitt[2] calculated the hole mass evaporation ratio, and Zurek[6] compared the decrease of the black-hole entropy with the increase of the entropy of its surroundings for massless particles. Page,[7] using black-hole perturbation methods of Teukolsky and Press,[8] has calculated the transmission coefficient for massless particles: neutrinos, photons, and gravitons (and possibly ultrarelativistic electrons and positrons for a small enough hole). Analytic expressions were given for the limiting cases $M\omega \gg 1$ and $M\omega \ll 1$. Massless particles will dominate the emission when $M \gg 10^{17}$ g. If 5×10^{14} g $< M < 9.4\times 10^{16}$ g, electrons and positrons are emitted, and a hole with $M \leqslant 5\times 10^{14}$ g, would emit muons and heavier particles at a significant rate. The properties of a Schwarzschild black hole as an elastic scatterer of waves were studied in detail by Sanchez.[9–12]

For a Schwarzschild black hole the emission and absorption spectra are related by[1]

$$H(\omega) = \sigma(\omega)[\exp(8\pi M\omega) \pm 1]^{-1}, \tag{2}$$

where $H(\omega)$ and $\sigma(\omega)$ stand for the emission and the absorption rates, respectively.

We have calculated the transmission coefficient and $H(\omega)$ and $\sigma(\omega)$ of an uncharged, nonrotating black hole for scalar particles of rest mass μ in a mode of energy ω, and angular momentum l, using the JWKB approximation. In the adiabatic semiclassical approximation, the black hole is assumed to have a classical Schwarzschild metric with $M \gg 1 \cong 2\times 10^{-5}$ g. We use Jacobian elliptic functions and integrals to obtain an analytic expression for the barrier penetration. These functions have been successfully applied to the resolution of other barrier penetration problems of nonlinear symmetrical and asymmetrical potentials.[13]

II. TRANSMISSION COEFFICIENT

To simplify the notation, the energy is denoted as a reduced energy $\widetilde{E} = \omega/\mu$. The effective potential $\widetilde{V}$ (also in units of μ) created by a black hole, in the Schwarzschild geometry, has a barrier of finite thickness that depends on the hole's mass and angular momentum. A particle created by a hole can escape by passing over, or by tunneling through, the barrier. The reduced energy must be greater than or equal to unity. If $\widetilde{E} < 1$ and the particle has managed to cross the barrier, then it will stay orbiting around the hole. In this paper we calculate the absorption in the region $\widetilde{V}_M \geqslant \widetilde{E} > 1$ or $\widetilde{V}_M \geqslant \omega > \mu$. If $\widetilde{E} > \widetilde{V}_M$ we take $\Gamma \simeq 1$.

Taking into account the quantization of angular momentum, the gravitational potential energy in the Schwarzschild metric is[14]

$$V(r) \simeq (1 - 2M/r)^{1/2}[\mu^2 + l(l+1)/r^2]^{1/2}, \tag{3}$$

where l is an eigenvalue of the angular momentum. The coordinate system r, ψ, and θ is chosen so that the radial projection of the orbit coincides with the equator, $\theta = \pi/2$.

The transmission coefficient, in first-order JWKB approximation, is

$$\Gamma = [1 + \exp(2\Phi)]^{-1}, \tag{4}$$

where

$$\Phi = \int_{r_2}^{r_3} (V^2 - \omega^2)^{1/2}\, dr \tag{5}$$

and the r_i are the roots of the equation

$$\omega^2 - V^2 = 0. \tag{6}$$

In Eq. (5) we assume $r_1 \leqslant r_2 \leqslant r_3$.

The energy conservation equation is[15]

$$\left\{\frac{[l(l+1)]^{1/2}}{r^2}\frac{dr}{d\psi}\right\}^2 + V^2(r,l) = \omega^2. \tag{7}$$

Substituting $u = M/r$, an analytic solution of Eq. (7) is achieved using the Jacobian elliptic funtions. In the region $u_3 \leqslant u \leqslant u_2$, the solution of Eq. (7) is[16]

$$u = R\,\mathrm{sn}^2(h^*\psi;m) + u_3, \tag{8}$$

 0022-2488/87/112665-04$02.50

where sn is the elliptic function senam with a parameter m.[17] The parameters R, h^*, and m depend on the roots r_i (u_i):

$$R = u_2 - u_1, \tag{9}$$

$$m = (u_2 - u_3)/(u_1 - u_3), \tag{10}$$

$$h^{*2} = -h^2 = (u_1 - u_3)/2. \tag{11}$$

Then Eq. (5) takes the form

$$\Phi = i[l(l+1)]^{1/2} \int_{u_2}^{u_3} \frac{1}{u^2} \left[\frac{du}{d\psi}\right] du, \tag{12}$$

where $i = (-1)^{1/2}$.

An analytic solution of Eq. (12) can also be obtained using the elliptic functions. Differentiating u with respect to ψ in Eq. (8) and substituting it in Eq. (12) one gets

$$\Phi = -\frac{8[l(l+1)]^{1/2} m^2 h^5}{u_3^2} \times \int_0^{K(m)} \frac{\mathrm{sn}^2 u \, \mathrm{cn}^2 u \, \mathrm{dn}^2 u}{(1 - \alpha^2 \mathrm{sn}^2 u)^2} du, \tag{13}$$

where $\alpha^2 = -R/u_3$ and $K(m)$ is the complete elliptic integral of the first kind. The only difficult calculation in Eq. (13) is the integral

$$[I(\alpha,m)]_0^1 \equiv \int_0^{K(m)} \frac{\mathrm{sn}^2 u \, \mathrm{cn}^2 u \, \mathrm{dn}^2 u}{(1 - \alpha^2 \mathrm{sn}^2 u)^2} du = \int_0^1 \frac{x^2 y}{(1 - \alpha^2 x^2)^2} dx, \tag{14}$$

where $x = \mathrm{sn}\, u$ and $y^2 = (1 - x^2)(1 - k^2 x^2)$. We have used two methods to carry out this calculation: first by using the Byrd and Friedman tables[18]; second, more generally and elegantly, using the reduction formulas given by Rodriguez Sanjuan.[19] The calculations corresponding to our case are to be found in the Appendix and give, with $m = k^2$, the complete integral

$$[I(\alpha,m)]_0^1 = -(3/2\alpha^4)E(m) + (1/2\alpha^6)[3m + (1-2m)\alpha^2]K(m) - (1/2\alpha^6)[\alpha^4 - 2(1+m)\alpha^2 + 3m]\Pi(\alpha^2,m), \tag{15}$$

where $E(m)$, $K(m)$, and $\Pi(\alpha^2,m)$ are the complete elliptic integrals of second, first, and third kind.

One finding is that Γ depends on the product $M\mu$. The cases $M\mu = 0.0019$ and $M\mu = 0.043$ are important. The first would correspond to a black hole of mass $M = 4.587 \times 10^{19} = 10^{15}$ g that would emit particles of mass $\mu = 4.19 \times 10^{-25} = 9.1 \times 10^{-28}$ g (electron and positron mass). The second would correspond to a hole of mass $M = 4.2 \times 10^{18} = 8.5 \times 10^{13}$ g that would emit particles of mass $\mu = 1.1 \times 10^{-20} = 2.4 \times 10^{-25}$ g (pion mass).

The dependence of Φ on the angular momentum is shown in Fig. 1 for several values of the reduced energy. There is a rapid increase of Φ with l. Therefore the transmission coefficient is important for only one value of l if $\tilde{V}_M > \tilde{E} > 1$, because of the exponential relationship between Γ and Φ. Thus with $M\mu = 0.0019$ and $\tilde{E} = 200$, $\Gamma \simeq 1$ for $l = 1$ $(\tilde{E} > \tilde{V}_M)$, $\Gamma = 6.8 \times 10^{-6}$ for $l = 2$, and $\Phi = 362$ and $\Gamma \simeq 0$ for $l = 3$.

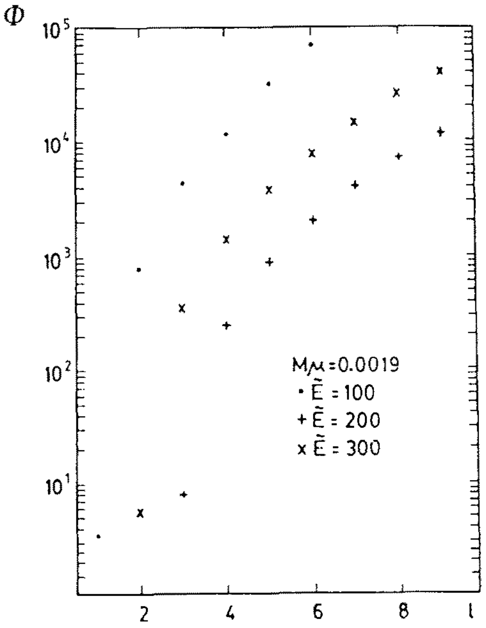

FIG. 1. The coefficient Φ as a function of the angular momentum eigenvalues is given for the reduced energy values $\tilde{E} = 100$, 200, and 300 when the mass product of the black hole and the particles is $M\mu = 0.0019$.

The value of Γ increases nearly exponentially with the reduced energy $\tilde{E}$, for fixed values of l and $M\mu$. When $\tilde{E} = \tilde{V}_M$ then $\Gamma = \frac{1}{2}$ (Fig. 2). One can obtain the Γ vs $\tilde{E}$ curves for $l = 3,5,...$ by the translation of the $l = 1$ curve.

Several values of the transmission coefficient as a function of the reduced energy and the mass product $M\mu$ are given in Table I. The angular momentum and barrier maximum are also indicated. There is the same value of Γ for particles of small mass emitted by a hole of big mass as for heavy particles emitted by a hole of small mass. However, Γ grows as the inverse of the product $M\mu$. Therefore $M\mu$ has a lower bound due to the quantization of the angular momentum. For example, if $\tilde{E} = 1.046$ then $M\mu = 1$ is the minimum value possible, and if $\tilde{E} = 19.3$ then it is $M\mu = 0.01$.

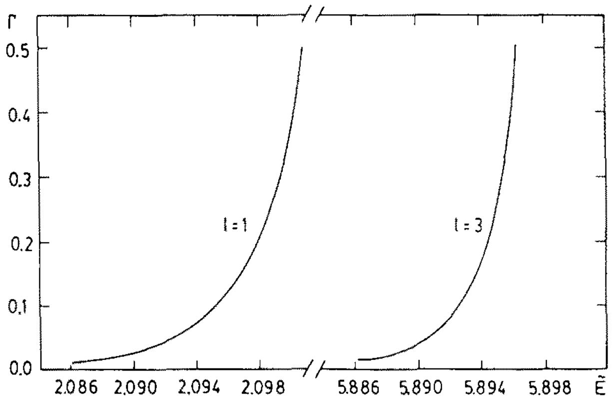

FIG. 2. Transmission coefficient Γ as a function of the reduced energy $\tilde{E}$ when the mass product $M\mu$ is 0.01 and the angular momentum values are $l = 1$ and $l = 3$.

TABLE I. The transmission coefficient Γ is given for several values of the reduced energy $\tilde{E} = \omega/\mu$, the angular momentum, and the mass product $M\mu$ of the black hole and the particle.

$\tilde{E}$	$M\mu$	l	Γ
	0.1	2	5.3×10^{-3}
3.970	1	20	8.9×10^{-20}
	10	200	~0
	0.1	2	0.068
3.980	1	20	1.7×10^{-10}
	10	200	3.7×10^{-96}
	0.1	3	0.013
5.895	1	30	2.1×10^{-17}
	10	300	~0
	0.1	3	0.33
5.896	1	30	6.1×10^{-5}
	10	300	3.1×10^{-42}
	0.01	1	0.1
	0.1	10	5.3×10^{-8}
19.300	1	100	1.7×10^{-70}
	10	10^3	~0

III. ABSORPTION CROSS SECTION

The absorption cross section for particles with mass in the Schwarzschild geometry is

$$\sigma(\tilde{E}) = \frac{\Sigma_l (2l+1)\Gamma}{\mu^2(\tilde{E}^2 - 1)}. \qquad (16)$$

We have calculated $\sigma(\tilde{E})$ for a wide range of values of the reduced energy, with the values of Γ obtained from Eq. (4) (Fig. 3), and it is seen to oscillate around its geometrical optics value $27\pi M^2$ as in the case of an elastic scatterer of waves.[11] These oscillations have a period equal to the value of the position of the potential maximum. The difference between maximum and minimum gets smaller as the reduced energy increases. In the limit $\sigma(\infty) \to 27\pi M^2$.

When the reduced energy is sufficiently small (for example, $\tilde{E} \lesssim 4.3$ for the case $M = 10^{14}$ g and $\mu = 1.9\times10^{-25}$ g) the total absorption cross section tends to zero. This is because the width of the potential barrier is big for all values of the angular momentum, and therefore $\Gamma \simeq 0$.

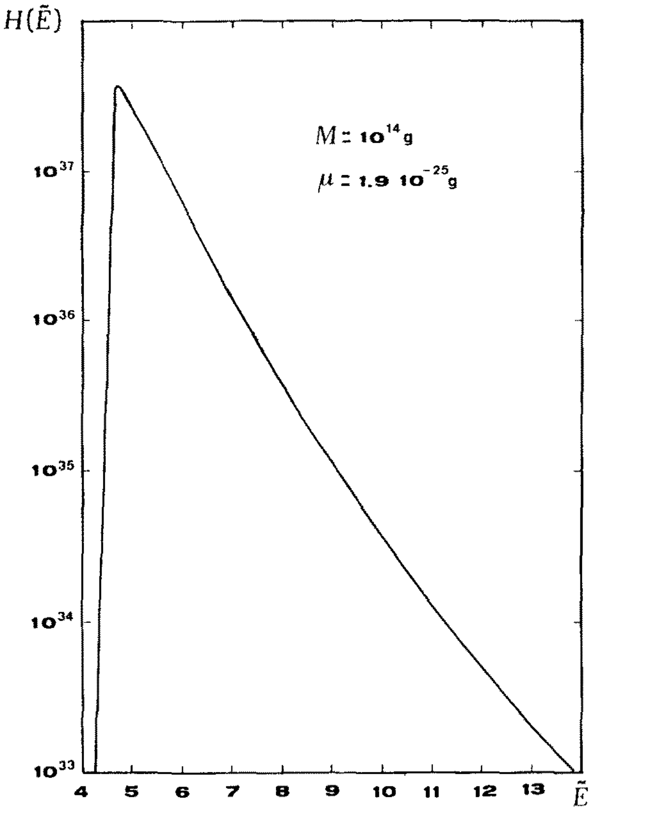

FIG. 4. Emission rates H as a function of the reduced energy for a black hole of 10^{14} g that evaporates particles of the muon mass.

IV. HAWKING EMISSION

In the context of classical field theory, black holes absorb particles but they cannot emit them. However, if quantum effects are considered, the absorption and emission spectrum are related by Eq. (2).

In Fig. 4 we plot $H(\tilde{E})$ as a function of the reduced energy. We see that it does not show any of the oscillations characteristic of the total absorption cross section (Fig. 3). This is due to the rapid decrease of the Planck factor for $M\omega = M\mu\tilde{E} \gtrsim 1$. Hawking emission is only significant in the energy range $1 \leqslant \tilde{E} \lesssim 1/M\mu$. These results are similar to the results of Sanchez for waves.[12]

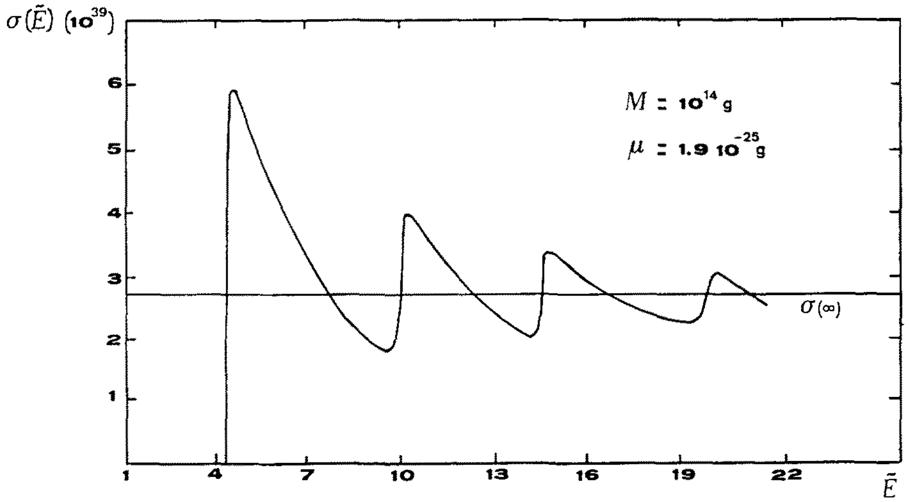

FIG. 3. Absorption cross section σ as a function of the reduced energy for a black hole of 10^{14} g that absorbs particles of the muon mass.

V. CONCLUSIONS

When we calculate the transmission coefficient Γ and the absorption cross section σ of a black hole in the Schwarzschild metric for scalar particles with mass, we find that Γ depends on the mass product of the particle and the hole: the smaller $M\mu$, the bigger Γ. Thus there is a lower limit to $M\mu$ due to the quantization of the angular momentum.

The absorption cross section as a function of the reduced energy $\widetilde{E}$ oscillates, with a period equal to the value of the position of the potential maximum. The difference between maximum and minimum gets smaller as the reduced energy increase and $\sigma(\infty) \to 27\pi M^2$.

The emission spectrum does not show any of the oscillations characteristic of the absorption cross section. This is related to the rapid decrease of the Planck factor for $\widetilde{E} \gtrsim 1/M\mu$. Hawking emission is only important in the energy range $1 \ll \widetilde{E} \lesssim 1/M\mu$ or $\mu \ll \omega \lesssim 1/M$.

ACKNOWLEDGMENT

The authors wish to thank A. Rodriguez Sanjuan for giving us formulas (A1) and (A4), and also for the details of the calculation.

APPENDIX: RODRIGUEZ SANJUAN INTEGRATION METHOD

For the calculation of integral I from Eq. (14) we follow a method of Rodriguez Sanjuan[19] for an extensive group of elliptic integrals. In our case this calculation is also quite simple. With the ansatz

$$I = \frac{Axy}{1-\alpha^2x^2} + \int \frac{Bx^2+C}{y}\,dx + \int \frac{D}{(1-\alpha^2x^2)y}\,dx, \tag{A1}$$

derivation of both sides gives

$$\begin{aligned}\frac{x^2y}{(1-\alpha^2x^2)^2} &= A\left[\frac{1+\alpha^2x^2}{(1-\alpha^2x^2)^2}\,y - \frac{x^2}{y}\,\frac{1+k^2-2k^2x^2}{1-\alpha^2x^2}\right] \\ &\quad + \frac{Bx^2+C}{y} + \frac{D}{(1-\alpha^2x^2)y}.\end{aligned} \tag{A2}$$

To determine A and D, set $x^2 = 1/\alpha^2$ in (A2). Then $A = 1/2\alpha^2$ and

$$x^2 - A(1+\alpha^2x^2) = -(1-\alpha^2x^2)/2\alpha^2$$

from which follows

$$\begin{aligned}(1/2\alpha^2)(1-x^2)(1-k^2x^2) &= (x^2/2\alpha^2)(1+k^2-2k^2x^2) - D \\ &\quad - (Bx^2+C)(1-\alpha^2x^2).\end{aligned} \tag{A3}$$

With $x^2 = 1/\alpha^2$ in (A3),

$$\frac{1}{2\alpha^2}\left(1-\frac{1}{\alpha^2}\right)\left(1-\frac{k^2}{\alpha^2}\right) = \frac{1}{2\alpha^4}\left(1+k^2-\frac{2k^2}{\alpha^2}\right) - D,$$

i.e.,

$$D = (1/2\alpha^6)[\alpha^4 - 2(1+k^2)\alpha^2 + 3k^2].$$

Setting $x^2 = 0$ in (A3), then

$$C = (1/2\alpha^6)[3k^2 - 2(1+k^2)\alpha^2].$$

The coefficient of x^4 gives $B = 3k^2/2\alpha^4$, and consequently

$$\begin{aligned}\int \frac{Bx^2}{y}\,dx &= -\frac{3}{2\alpha^4}\int \frac{(1-k^2x^2)-1}{y}\,dx \\ &= -\frac{3}{2\alpha^4}(E-F),\end{aligned}$$

where E and F are the incomplete elliptic integrals of the second and first kinds. Then finally we have

$$\begin{aligned}I &= \frac{1}{2\alpha^2}\,\frac{xy}{1-\alpha^2x^2} - \frac{3}{2\alpha^4}E(x,k^2) \\ &\quad + \frac{1}{2\alpha^6}[(1-2k^2)\alpha^2 + 3k^2]F(x,k^2) \\ &\quad - \frac{1}{2\alpha^6}[\alpha^4 - 2(1+k^2)\alpha^2 + 3k^2]\Pi(x,\alpha^2,k^2).\end{aligned} \tag{A4}$$

[1]S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975).
[2]B. S. DeWitt, Phys. Rep. C **19**, 295 (1975).
[3]R. Balbinot, R. Bergami, and B. Giorgini, Nuovo Cimento B **71**, 27 (1982).
[4]M. I. Beciu, Phys. Lett. A **100**, 77 (1984).
[5]Y. Kuroda, Prog. Theor. Phys. **71**, 100 (1984).
[6]W. H. Zurek, Phys. Lett. A **77**, 339 (1980).
[7]D. N. Page, Phys. Rev. D **13**, 198 (1976).
[8]S. A. Teukolsky and W. H. Press, Astrophys. J. **193**, 443 (1974).
[9]N. G. Sanchez, J. Math. Phys. **17**, 688 (1976).
[10]N. G. Sanchez, Phys. Rev. D **16**, 937 (1977).
[11]N. G. Sanchez, Phys. Rev. D **18**, 1030 (1978).
[12]N. G. Sanchez, Phys. Rev. D **18**, 1798 (1978).
[13]J. Díaz, A. Martín, and C. Miró, J. Chem. Phys. **85**, 5128 (1986).
[14]N. Deruelle and R. Ruffini, Phys. Lett. B **52**, 437 (1974).
[15]S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford U. P., London, 1983).
[16]P. A. Collins, R. Delbourgo, and R. M. Williams, J. Phys. A: Math. Gen. **6**, 161 (1973).
[17]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1972).
[18]P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Scientists* (Springer, Berlin, 1971).
[19]A. Rodriguez Sanjuan, Rev. Real Acad. Ciencias **50**, 1 (1956).

Nonorientable one-loop amplitudes for the bosonic open string: Electrostatics on a Möbius strip

João P. Rodrigues
Physics Department, University of the Witwatersrand, Johannesburg 2001, South Africa

(Received 16 September 1986; accepted for publication 10 June 1987)

The partition function, N-point scalar, and four-point vector nonorientable one-loop amplitudes for the bosonic open string in the critical dimension are obtained using a first quantized path integral treatment of Polyakov's string that assumes scale independence.

I. INTRODUCTION

Numerous efforts are now concentrated on the construction of a second quantized, field theory of strings.[1] Such a formalism will hopefully enable one to understand, amongst other issues, the origin of general covariance and the nontriviality of the vacuum structure, and its importance cannot be overestimated. At this stage, however, this program is still incomplete.

Another approach to string theories is to consider them as two-dimensional models whose basic variables are the string coordinates x_μ, $\mu = 1,2,...,d$, $\sigma = (\sigma^1,\sigma^2)$. Hsue, Sakita, and Virasoro[2] were the first to use this first quantized approach to obtain tree and one-loop N-point tachyon amplitudes in the old dual resonance model (bosonic open string). A closer look at their paper, however, shows that the partition function for the one-loop amplitudes was still computed with operator methods.

Polyakov's path integral formulation of the bosonic string[3] provided the major step in the understanding of the correct string functional integral measure and was the basis of much further work (see for instance Refs. 4 and 5).

With the recent interest in string theories, Polyakov's string theory was used as a starting point in the first quantized evaluation of N-point covariant tree amplitudes for the bosonic string in the critical dimension.[6]

However, one-loop amplitudes had to wait for a fuller understanding of the functional integral measure, particularly in that the question of moduli and the nature of Teichmuller space are concerned. There has been considerable recent progress on these questions,[7–9] and in particular a one-loop analysis of the closed bosonic string has been carried out.[7] The question of fermionic degrees of freedom has also been addressed.[10] One is then in position to obtain covariant amplitudes, such as a string–string propagator, as results that a second quantized formalism must reproduce. This has already been achieved in Ref. 11.

There are other aspects, however, for which the first quantized approach seems particularly suited, such as the possibility of using string theories to obtain the effective action of gauge fields.[12,13] For this, one needs to consider the open string, and in particular vector amplitudes, if one wants to go beyond the string level contribution. Moreover, it is likely that multiloop amplitudes will be first obtained using this approach, opening the way to discussions of higher-loop finiteness in the superstring case.

I have recently presented a first quantized derivation of planar and nonplanar one-loop amplitudes[14] for the open bosonic string. The aim of this communication is to generalize the approach to the nonorientable case. To the best of my knowledge, a first quantized derivation of the nonorientable amplitude does not exist in the literature.

The method used is that of Polchinski's,[7] where scale invariance is assumed at every step. Moore and Nelson[8] have given it a sound theoretical foundation and I refer the reader to their paper, where the problem of scale invariance and moduli is clearly presented and solved.

The question of boundaries in string theory has been considered by Alvarez,[5] and Neumann boundary conditions were considered in Ref. 14. In his work, Alvarez[5] always dealt with orientable topologies. Although it is not my intention to generalize the method to general nonorientable topologies (since its validity to the particular case considered in this paper is always obvious), it is important to find that it reproduces the scalar amplitudes obtained with operator methods. More importantly, the overall normalization is uniquely fixed. The relative normalization between the nonorientable and the planar amplitude is of course crucial, for the superstring, in the establishment of one-loop finiteness.[15] Furthermore, vector amplitudes are derived.

In the next section the partition function is obtained. In Sec. III, the N-point scalar amplitude is computed and compared to known one-loop results. This serves as a check of the result of Sec. II. The four-point vector amplitude will be presented in Sec. IV.

II. PARTITION FUNCTION

The starting point is Polyakov's path integral,[3]

$$W = \int \frac{dg_{ab}\, dx^\mu}{V_{GC} V_W} \exp\biggl(- \int d^2\sigma \sqrt{g} \times \left[\frac{1}{4\pi\alpha} g^{ab}\, \partial_a x^\mu\, \partial_b x^\mu + \lambda R + \mu^2 \right] + S_b \biggr). \quad (1)$$

One integrates over all Euclidean metrics $g_{ab}(\sigma)$ on a surface of fixed topology and all embeddings $x^\mu(\sigma)$ into R^d. Here, R is the scalar curvature. The action is invariant under general coordinate invariance of the world sheet and classically also under Weyl transformations:

$$\delta g_{ab}(\sigma) = \lambda(\sigma) g_{ab}(\sigma). \quad (2)$$

Polyakov's analysis shows that this symmetry is broken at the quantum level but that if $d = 26$ then μ^2 can be chosen

 0022-2488/87/112669-08$02.50

in such a way as to regain scale invariance at the full quantum level. Because of the presence of boundaries, further terms are required to maintain quantum Weyl invariance.[5] These are contained in S_b, which will be discussed later.

The topology for the nonorientable one-loop open string is that of a Möbius strip described by the primitive cell

$$0 \leqslant \text{Im}\,\sigma \leqslant 1, \quad 0 \leqslant \text{Re}\,\sigma \leqslant 1, \tag{3}$$

with the points (1,0) and (1,1) identified with (0,1) and (0,0), respectively, in the usual manner.

Since it is a world sheet scalar, $x^\mu(\sigma)$ satisfies the "twisted" boundary condition

$$x^\mu(\sigma^1+1,\sigma^2) = x^\mu(\sigma^1,1-\sigma^2). \tag{4a}$$

World sheet vectors obey

$$d\sigma^1 \to d\sigma^1, \quad \partial_1 \to \partial_1$$
$$\text{under } (\sigma^1+1,\sigma^2) \to (\sigma^1,1-\sigma^2),$$
$$d\sigma^2 \to -d\sigma^2, \quad \partial_2 \to -\partial_2 \tag{4b}$$
$$\text{under } (\sigma^1+1,\sigma^2) \to (\sigma^1,1-\sigma^2).$$

This is a result of the fact that the direction of the normal changes in going from the $\sigma^2 = 0$ to the $\sigma^2 = 1$ boundary. It follows that the metric tensor satisfies

$$g_{11}(\sigma^1+1,\sigma^2) = g_{11}(\sigma^1,1-\sigma^2),$$
$$g_{12}(\sigma^1+1,\sigma^2) = -g_{12}(\sigma^1,1-\sigma^2), \tag{4c}$$
$$g_{22}(\sigma^1+1,\sigma^2) = g_{22}(\sigma^1,1-\sigma^2).$$

For the open string, Neumman's boundary conditions must be imposed for $x^\mu(\sigma)$, i.e.,

$$\frac{\partial x^\mu}{\partial\sigma^2}(\sigma^2=0) = \frac{\partial x^\mu}{\partial\sigma^2}(\sigma^2=1) = 0. \tag{5a}$$

It should be remembered that this is a sufficient condition for the existence of a classical extremum of the action.[5]

The condition to be imposed on the metric is the following[11,14]: If t^a is the tangent vector to the boundary and n^a an arbitrary normal vector to the boundary, it is required that

$$t^a n^b g_{ab} = 0 \quad \text{on the boundary.} \tag{5b}$$

This requirement is independent of the choice of n^a, and therefore it does not provide any new geometric information.[11] However, it is important in that it provides a well-defined mode expansion for the metric and metric variations.[5]

By a general coordinate transformation, any metric can be transformed, while satisfying (4), to

$$ds^2 = g_{ab}\,d\sigma^a\,d\sigma^b = e^{\phi(\sigma)}[(d\sigma^1)^2 + \tau^2(d\sigma^2)^2], \tag{6}$$

which corresponds to a metric proportional to the Euclidean metric on the rectangle of sides 1 and $i\tau$ (a Möbius strip of width τ). For convenience, and following Ref. 7, I keep the unit cell to the square $0 \leqslant \sigma^a \leqslant 1$, $a = 1,2$.

The positive real number τ is the modulus of the Möbius strip and it parametrizes the one-dimensional Teichmuller space of the topology.

Any variation of the metric connected to unity can be decomposed as

$$dg_{ab}(\sigma) = g_{ab}(\sigma)\delta\phi(\sigma) + \delta\zeta_{a;b}(\sigma) + \delta\zeta_{b;a}(\sigma) + g_{ab,\tau}d\tau. \tag{7}$$

The method of Ref. 7 consists in finding the Jacobian $J(\phi,\tau)$ defined by

$$dg = (d\phi\,d\zeta)'\,d\tau\,J(\phi,\tau), \tag{8}$$

where the prime denotes variations orthogonal to the zero translational mode (conformal Killing vector)

$$\delta\zeta^1(\sigma) = \epsilon^1, \quad \delta\phi(\sigma) = -\epsilon^1\,\partial_1\phi(\sigma). \tag{9}$$

Metrics for small changes in the fields are defined in the usual way:

$$\|\delta g\|^2 = \int d^2\sigma\sqrt{g}(g^{ac}g^{bd} + Cg^{ab}g^{cd})\delta g_{ab}\,\delta g_{cd},$$
$$\|\delta\phi\|^2 = \int d^2\sigma\sqrt{g}\,\delta\phi^2, \tag{10}$$
$$\|\delta\zeta\|^2 = \int d^2\sigma\sqrt{g}\,g^{ab}\,\delta\zeta_a\,\delta\zeta_b,$$
$$\|\delta\mathbf{x}\|^2 = \int d^2\sigma\sqrt{g}\,\delta x^\mu\,\delta x^\mu,$$

where C is arbitrary. Following Ref. 7, the measure is defined implicitly by the following general expression:

$$\int d\delta\psi e^{-(1/2)\|\delta\psi\|^2} = 1, \tag{11a}$$

where $\delta\psi = \delta g$, $\delta\phi$, $\delta\zeta$, or $\delta\mathbf{x}$. Also,

$$\int d\delta\,\tau\exp\left(-\frac{1}{2}\delta\tau^2\int d^2\sigma\sqrt{g}\right) = \left(\frac{2\pi}{\int d^2\sigma\sqrt{g}}\right)^{1/2}. \tag{11b}$$

A careful analysis of why this is a correct procedure has been given in Ref. 8.

Separating the zero mode from the integrals involving $\delta\phi$ and $\delta\zeta$, I obtain

$$1 = \sqrt{\frac{2\pi}{Q}}\int(d\delta\phi\,d\delta\zeta)'e^{-(1/2)(\|\delta\phi\|^2 + \|\delta\zeta\|^2)} \tag{12}$$

where $Q = \int d^2\sigma\sqrt{g}(\partial_1\phi\,\partial_1\phi + 1)$. Then the integral involving $d\delta g$ becomes

$$1 = J(\phi,\tau)\int(d\delta\phi\,d\delta\zeta)'\,d\delta\tau\,e^{-(1/2)\|\delta g\|^2}, \tag{13}$$

with

$$\|\delta g\|^2 = \int d^2\sigma\sqrt{g}(\delta\phi \quad \delta\zeta^a \quad \delta\tau)\mathscr{M}\begin{pmatrix}\delta\phi\\ \delta\zeta_b\\ \delta\tau\end{pmatrix}. \tag{14}$$

With the hindsight of Ref. 7, the matrix $\mathscr{M}$ can be written as

$$\mathscr{M} = \mathscr{A}\mathscr{N}\mathscr{A}^T = \begin{pmatrix}1 & 0 & 0\\ -D_a & \delta_a^c & 0\\ \frac{1}{2}g^{ef}g_{ef,\tau} & 0 & 1\end{pmatrix}\begin{pmatrix}2+4K & 0 & 0\\ 0 & 2\Delta_c^d & -2D_eX_c^e\\ 0 & 2X^{ed}D_e & X_{ef}X^{ef}\end{pmatrix}\begin{pmatrix}1 & D^b & \frac{1}{2}g^{ef}g_{ef,\tau}\\ 0 & \delta_d^b & 0\\ 0 & 0 & 1\end{pmatrix}, \tag{15}$$

where $\Delta_c^d = -\delta_c^d D^2 - D^d D_c + D_c D^d$ and $X_{ab} = g_{ab,\tau} - \frac{1}{2} g_{ab}\, g^{cd} g_{cd,\tau}$.

It then follows that

$$J(\phi,\tau) = (\det{}' \mathcal{N})^{1/2} \left(\frac{\int d^2\sigma \sqrt{g}}{Q} \right)^{1/2}. \tag{16}$$

The x^μ integration is carried out in the usual way by shifting from the constant configuration [recall that (5a) ensures that the action has a classical extremum], and once use is made of (11a) one obtains

$$\int d\mathbf{x} \exp\left(-\frac{1}{4\pi\alpha} \int d^2\sigma \sqrt{g} g^{ab}\, \partial_a x^\mu\, \partial_b x^\mu \right) = L^d \left(\frac{2\pi}{\int d^2\sigma \sqrt{g}} \det{}' \left(-\frac{1}{2\pi\alpha} \frac{1}{\sqrt{g}} \partial_a\, g^{ab} \sqrt{g}\, \partial_b \right) \right)^{-d/2}. \tag{17}$$

The system has been put in a hypercube of side L.

The partition function of the Möbius strip then takes the form

$$W_{\text{Möbius}} = \int \frac{(d\xi\, d\phi)'}{V_{GC} V_W} L^d (\det{}' \mathcal{N})^{1/2} \left(\int d^2\sigma \sqrt{g} \right)^{(d+1)/2} Q^{-1/2} (2\pi)^{-d/2} \left[\det{}' \left(-\frac{1}{2\pi\alpha} \frac{1}{\sqrt{g}} \partial_a\, g^{ab} \sqrt{g}\, \partial_b \right) \right]^{-d/2}. \tag{18}$$

It is well known[3] that the bulk ϕ dependence in the above expression can be made to vanish in 26 dimensions by a suitable adjustment of μ^2. It was explicitly shown in Ref. 11 what further local, boundary-dependent terms S_b are required so that all scale dependence is renormalized away. One can then set $\phi = 0$.

The condition (5b) and the fact that the region must remain unchanged imply, for the metric variations (7), that

$$\partial_2 \xi^1(\sigma^2 = 0) = \partial_2 \xi^1(\sigma^2 = 1) = 0, \tag{19a}$$

$$\xi^2(\sigma^2 = 0) = \xi^2(\sigma^2 = 1) = 0. \tag{19b}$$

It must be remembered that ξ_1 and ξ_2 obey different "twisted" boundary conditions [Eq. (4b)]. Since $\int d^2\sigma \sqrt{g} = \tau$, $Q = \tau$, one obtains

$$\begin{aligned}(\det{}' \mathcal{N})^{1/2} &= [\det(2 + 4C)]^{1/2} (\sqrt{2}/\tau) [\det{}'(-2\delta_c^d\, g^{ab}\, \partial_a\, \partial_b)]^{1/2} \\ &= [\det(2 + 4C)]^{1/2} (\sqrt{2}/\tau) [\det{}'_{\mathrm{N}}(-2g^{ab}\, \partial_a\, \partial_b)]^{1/2} [\det_{\mathrm{D}}(-2g^{ab}\, \partial_a\, \partial_b)]^{1/2} \\ &= [\det(2 + 4C)]^{1/2} [\det_{\mathrm{N}}(2)]^{1/2} [\det_{\mathrm{D}}(2)]^{1/2} (1/\tau) [\det{}'_{\mathrm{N}}(-\Delta)]^{1/2} [\det_{\mathrm{D}}(-\Delta)]^{1/2},\end{aligned} \tag{20}$$

$$\det{}' \left(-\frac{1}{2\pi\alpha} \sqrt{g}^{-1}\, \partial_a\, g^{ab} \sqrt{g}\, \partial_b \right) = \det{}'_{\mathrm{N}} \left(-\frac{1}{2\pi\alpha} \Delta \right) = (2\pi\alpha) \det_{\mathrm{N}} \left(\frac{1}{2\pi\alpha} \right) \det{}'_{\mathrm{N}}(-\Delta), \tag{21}$$

where Δ is the scalar Laplacian and the subscripts D and N refer to Dirichelet (19b) and Neumann [(5a) and (19a)] boundary conditions, respectively.

The fields satisfying Neumann boundary conditions also satisfy periodic "twisted" boundary conditions [Eqs. (4a) and (4b)]. A basis for such functions is the set

$$e^{i\pi n_1 \sigma^1} \cos(n_2 \pi \sigma^2), \quad n_1, n_2 \geqslant 0, \quad \text{both even, both odd.} \tag{22}$$

The vector ξ_2, satisfying Dirichelet boundary conditions [Eq. (19b)], obeys antiperiodic "twisted" boundary conditions [Eq. (4b)]. A basis for such functions is the set

$$e^{i\pi n_1 \sigma^1} \sin(n_2 \pi \sigma^2), \quad n_1, n_2 > 0, \quad \text{both even, both odd.} \tag{23}$$

Therefore

$$\det{}'_{\mathrm{N}}(-\Delta) = \prod_{\substack{n_1\ \text{even} \\ n_2 \geqslant 0\ \text{even}}}{}' \left[\pi^2 n_1{}^2 + \frac{\pi^2 n_2^2}{\tau^2} \right] \times \prod_{\substack{n_1\ \text{odd} \\ n_2 \geqslant 0\ \text{odd}}} \left[\pi^2 n_1{}^2 + \frac{\pi^2 n_2^2}{\tau^2} \right],$$

$$\det_{\mathrm{D}}(-\Delta) = \prod_{\substack{n_1\ \text{even} \\ n_2 > 0\ \text{even}}} \left[\pi^2 n_1{}^2 + \frac{\pi^2 n_2^2}{\tau^2} \right] \times \prod_{\substack{n_1\ \text{odd} \\ n_2 > 0\ \text{odd}}} \left[\pi^2 n_1{}^2 + \frac{\pi^2 n_2^2}{\tau^2} \right]. \tag{24}$$

The prime indicates that the term $n_1 = n_2 = 0$ in the product must be removed. Then

$$\det_{\mathrm{D}}(-\Delta) \det{}'_{\mathrm{N}}(-\Delta) = \prod_{\substack{n_1\ \text{even} \\ n_2\ \text{even}}}{}' \left[\pi^2 n_1{}^2 + \frac{\pi^2 n_2^2}{\tau^2} \right] \prod_{\substack{n_1\ \text{odd} \\ n_2\ \text{odd}}} \left[\pi^2 n_1{}^2 + \frac{\pi^2 n_2^2}{\tau^2} \right], \tag{25}$$

$$\det{}'_{\mathrm{N}}(-\Delta) = \prod{}' [(2\pi)^2 n_1{}^2]^{1/2} \prod_{\substack{n_1\ \text{even} \\ n_2\ \text{even}}}{}' \left[\pi^2 n_1{}^2 + \frac{\pi^2 n_2^2}{\tau^2} \right]^{1/2} \times \prod_{\substack{n_1\ \text{odd} \\ n_2\ \text{odd}}} \left[\pi^2 n_1{}^2 + \frac{\pi^2 n_2^2}{\tau^2} \right]^{1/2}. \tag{26}$$

The τ-dependent products are evaluated in Appendix A with zeta function regularization and yield

$$\prod_{\substack{n_1\ \text{even}\\ n_2\ \text{even}}}{}' \left[\pi^2 n_1{}^2 + \frac{\pi^2 n_2^2}{\tau^2}\right]^{1/2}$$

$$= \tau e^{-\pi\tau/6}\left[\prod_{n=1}^{\infty}(1-e^{-2\pi n\tau})\right]^2, \qquad (27)$$

$$\prod_{\substack{n_1\ \text{odd}\\ n_2\ \text{odd}}} \left[\pi^2 n_1{}^2 + \frac{\pi^2 n_2^2}{\tau^2}\right]^{1/2}$$

$$= e^{\pi\tau/12}\left[\prod_{\substack{n=1\\ n\ \text{odd}}}^{\infty}(1+e^{-\pi n\tau})\right]^2. \qquad (28)$$

The τ-independent product in (26) is exactly 1 if zeta function regularization is used.[14] Collecting all factors together in Eq. (18), and neglecting all determinant factors[7] I obtain

$$W_{\text{Möbius}} = \frac{L^d}{2}\frac{1}{(4\pi^2\alpha)^{13}}\int_0^\infty d\tau \times e^{\pi\tau}\left[\prod_{n=1}^{\infty}(1-(-)^n e^{-\pi n\tau})\right]^{-24}. \qquad (29)$$

A factor of $\frac{1}{4}$ has been included in the above equation corresponding to the orientation preserving symmetry [respecting (4) and (5)] $\sigma^1 \to 1-\sigma^1$, $\sigma^2 \to 1-\sigma^2$ and the orientation changing transformation $\sigma^1 \to \sigma^1$, $\sigma^2 \to 1-\sigma^2$.

Transforming variables to

$$\tau = -(1/2\pi)\ln q, \qquad (30)$$

$W_{\text{Möbius}}$ becomes

$$W_{\text{Möbius}} = \frac{L^d}{8\pi}\frac{1}{(4\pi^2\alpha)^{13}}\int_0^1 dq \times q^{-3/2}\left[\prod_{n=1}^{\infty}(1-(-)^n\sqrt{q}^{\,n})\right]^{-24}. \qquad (31)$$

Notice that, as it is the case for the one-loop sum over surfaces for the closed string, the term containing λ in (1) does not contribute: the Euler characteristic is given by $\chi = 2-2h-b-c$, where h is the number of handles, b is the number of boundaries, and c is the number of cross caps yielding 0 in the present topology ($h=0$, $b=c=1$).

III. SCALAR AMPLITUDES

The N-point scalar (tachyon) amplitudes are well known to be generated by[12,13]

$$\left\langle \exp\left(g_0\int_{\partial M} ds\, A(x^\mu(s))\right)\right\rangle, \qquad (32)$$

where ∂M is the boundary and ds is the invariant line element (in our case, $ds = d\sigma^1$ when $\phi=0$). Fourier transforming,

$$A(x^\mu) = \int \frac{d^{26}k}{(2\pi)^{26}} e^{ik_\mu x^\mu} A(k^\mu), \qquad (33)$$

and for the one-point amplitude one needs to evaluate

$$\int dx \exp\left(-\frac{\tau}{4\pi\alpha}\int d^2\sigma\, \partial^a x^\mu \partial_a x^\mu\right)\exp\left(i\sum_1^N k_i^\mu x^\mu(s_i)\right). \qquad (34)$$

Because Neumann boundary conditions are imposed, some care must be taken. This was discussed in Ref. 14. The result is the "naive one":

$$(2\pi)^{26}\delta\left(\sum_1^N k_i^\mu\right)\exp\left(\frac{\tau}{4\pi\alpha}\int d^2\sigma\, x_N{}^\mu(-\partial^a\partial_a)x_N{}^N\right) \times\int dx'^\mu \exp\left(-\frac{\tau}{4\pi\alpha}\int d^2\sigma\,\partial^a x'^\mu \partial_a x'^\mu\right), \qquad (35)$$

where the normal component $\partial_n x'^\mu$ vanishes at the boundary and $x_N{}^\mu$ is the solution of Poisson's equation.

The Green's function for our problem

$$(\tau/2\pi\alpha)(-\partial^a\partial_a)G(\sigma,\sigma') = \delta(\sigma-\sigma') \qquad (36)$$

is

$$G(\sigma,\sigma') = \frac{\alpha\pi\tau}{6} - \frac{\alpha\pi\tau}{2}|\sigma^2-\sigma'^2| - \frac{\alpha\pi\tau}{2}|\sigma^2+\sigma'^2|_2 + \frac{\alpha\pi\tau}{2}|\sigma^2-\sigma'^2|^2 + \frac{\alpha\pi\tau}{2}|\sigma^2+\sigma'^2|_2^2$$
$$-\alpha\ln|1-e^{i\pi(\sigma^1-\sigma'^1)}e^{-\pi\tau|\sigma^2-\sigma'^2|}||1-e^{i\pi(\sigma^1-\sigma'^1)}e^{-\pi\tau|\sigma^2+\sigma'^2|}|$$
$$-\alpha\ln\prod_{n=1}^{\infty}|1-(-)^n e^{i\pi(\sigma^1-\sigma'^1)}e^{-\pi\tau|\sigma^2-\sigma'^2+n|}||1-(-)^n e^{i\pi(\sigma^1-\sigma'^1)}e^{-\pi\tau|\sigma^2+\sigma'^2+n|}|$$
$$-\alpha\ln\prod_{n=1}^{\infty}|1-(-)^n e^{i\pi(\sigma^1-\sigma'^1)}e^{-\pi\tau|\sigma^2-\sigma'^2-n|}||1-(-)^n e^{i\pi(\sigma^1-\sigma'^1)}e^{-\pi\tau|\sigma^2+\sigma'^2-n|}|, \quad \text{for } \sigma\neq\sigma', \qquad (37a)$$

$$G(\sigma,\sigma') = \frac{\alpha\pi\tau}{6} - \frac{\alpha\pi\tau}{2}|2\sigma^2|_2 + \frac{\alpha\pi\tau}{2}|2\sigma^2|_2^2 + \frac{\alpha}{2t} + \alpha\ln\left(\frac{e^\gamma}{2\pi}\right) - \alpha\ln(1-e^{-\pi\tau 2\sigma^2})$$
$$-\alpha\ln\prod_{n=1}^{\infty}(1-(-)^n e^{-\pi\tau n})^2 - \alpha\ln\prod_{n=1}^{\infty}(1-(-)^n e^{-\pi\tau(|n+2\sigma^2|)})(1-(-)^n e^{-\pi\tau(|n-2\sigma^2|)}),$$

for $\sigma=\sigma'$ and $0<\sigma^2<1$, (37b)

$$G(\sigma,\sigma') = \frac{\alpha\pi\tau}{6} + \frac{\alpha}{t} + 2\alpha\ln\left(\frac{e^\gamma}{2\pi}\right) - 2\alpha\ln\prod_{n=1}^{\infty}(1-(-)^n e^{-\pi\tau n})^2, \quad \text{for } \sigma=\sigma' \text{ and } \sigma^2 = 0,1. \qquad (37c)$$

The above propagator has been obtained using heat kernel regularization, since it is not defined for $\sigma = \sigma'$. Its derivation is described in Appendix B. In Eq. (37) the original period of 2 in the imaginary direction must be used to make $|\sigma^2 + \sigma'^2|_2 \leqslant 1$ (cf. Appendix B).

Using Eqs. (32), (35), and (36) as well as the results of the previous section, I obtain for the N-point scalar amplitude

$$A(k_1,k_2,\ldots,k_3) = \delta\Big(\sum_i k_i\Big)\frac{1}{8\pi}\frac{1}{(\alpha)^{13}}\frac{1}{N!}\Big(\prod_{i=1}^{N} g_0 \int_0^1 d\sigma_i^1\Big) \times \int_0^1 dq\, q^{-3/2}\Big[\prod_{i=1}^{\infty}(1-(-)^n\sqrt{q^n})\Big]^{-24} \exp\Big(-\frac{1}{2}\sum_{i\neq j} k_i^\mu k_j^\mu G'_{ij}\Big), \tag{38}$$

where

$$G'_{ij} = G(\sigma_i,\sigma_j) - \tfrac{1}{2}G(\sigma_i,\sigma_i) - \tfrac{1}{2}G(\sigma_j,\sigma_j). \tag{39}$$

In agreement with Ref. 14, the renormalized coupling constant g_R is defined as

$$g_R = \lim_{t\to 0} 2g_0(t)e^{-1/2t-\gamma}. \tag{40}$$

Then, on mass shell ($k^2 = \alpha^{-1}$),

$$A(k_1,k_2,\ldots,k_N) = \delta\Big(\sum_i k_i\Big)\frac{g_R^N}{8\pi}\frac{1}{(\alpha)^{13}}\Big(\prod_{i=1}^{N-1}\int_0^1 \theta(\sigma_{i+1}^1 - \sigma_i^1)d\sigma_i^1\Big) \times \int_0^1 dq\, q^{-3/2}\Big[\prod_{i=1}^{\infty}(1-(-)^n\sqrt{q^n})\Big]^{-24}\prod_{i>j}(\Psi_P \text{ or } \Psi_{NP})^{2\alpha k_i^\mu k_j^\mu},$$

where

$$\Psi_P = \frac{2}{\pi}\sin\Big(\frac{\pi(\sigma_i^1 - \sigma_j)}{2}\Big)\prod_{n=1}^{\infty}\frac{1-2(-)^n\cos(\pi(\sigma_i^1-\sigma_j^1))\sqrt{q^n}+q^n}{(1-(-)^n\sqrt{q^n})^2},$$

$$\Psi_{NP} = \frac{2}{\pi}\cos\Big(\frac{\pi(\sigma_i^1 - \sigma_j)}{2}\Big)\prod_{n=1}^{\infty}\frac{1+2(-)^n\cos(\pi(\sigma_i^1-\sigma_j^1))\sqrt{q^n}+q^n}{(1-(-)^n\sqrt{q^n})^2}, \tag{41}$$

for two points on the same and opposite boundaries, respectively. Of course one has only one boundary, and it is simple to see that $\Psi_P(\sigma^1 - \sigma'^1 + 1) = \Psi_{NP}$. Therefore extending the limits of integration from [0,1] to [0,2] one obtains contributions to both amplitudes.

For $\alpha = \frac{1}{2}$, these results are in perfect agreement with those of Ref. 16.

IV. VECTOR AMPLITUDE

The N-point vector amplitudes are generated by[12,13]

$$\Big\langle \exp\Big(g_0\int_{\partial M} ds\,\frac{dx^\mu(s)}{ds}A^\mu(x^\nu(s))\Big)\Big\rangle. \tag{42}$$

For simplicity, I will consider the four-point planar amplitude. The generalization is straightforward and will be presented elsewhere.[17] Fourier transforming and using the propagator of the previous section, I obtain

$$A^{\mu_1\mu_2,\ldots,\mu_N}(k_1,k_2,\ldots,k_N) = \delta\Big(\sum_i k_i\Big)\frac{g_R{}^N}{8\pi}\frac{1}{(\alpha)^{13}}\Big(\prod_{i=1}^{N-1}\int_0^1\theta(\sigma_{i+1}^1-\sigma_i^1)d\sigma_i^1\Big) \times \int_0^1 dq\,q^{-3/2}\Big[\prod_{i=1}^{\infty}(1-(-)^n\sqrt{q^n})\Big]^{-24}\prod_{i>j}(\Psi_P)^{2\alpha k_i^\mu k_j^\mu} \times[(Q^{\mu_1\mu_2}Q^{\mu_3\mu_4}+\cdots)-(Q^{\mu_1\mu_2}P_3^{\mu_3}P_4^{\mu_4}+\cdots)+P_1^{\mu_1}P_2^{\mu_2}P_3^{\mu_3}P_4^{\mu_4}], \tag{43a}$$

where

$$Q^{\mu_i\mu_j} = -2\pi^2\alpha\delta_{\mu_i\mu_j}\left[\frac{1}{4\sin^2(\pi/2)(\sigma_i^1-\sigma_j^1)} - \sum_{n=1}^{\infty} 2(-)^n\sqrt{q}^n \frac{(1-(-)^n\sqrt{q}^n)^2 - 2(1+q^n)\sin^2(\pi/2)(\sigma_i^1-\sigma_j^1)}{(1-2(-)^n\cos[\pi(\sigma_i^1-\sigma_j^1)]\sqrt{q}^n+q^n)^2}\right],$$

$$P^{\mu_i} = -2\pi\alpha\sum_{i\neq j} k_j^{\mu_i}\left[\frac{\cot(\pi/2)(\sigma_i^1-\sigma_j^1)}{2} + \sum_{n=1}^{\infty}\frac{2(-)^n\sqrt{q}^n\sin(\pi(\sigma_i^1-\sigma_j^1))}{1-2(-)^n\cos(\pi(\sigma_i^1-\sigma_j^1))\sqrt{q}^n+q^n}\right]. \qquad (43b)$$

In Eq. (43b), I have defined

$$\frac{d}{d\sigma_i^1}G'(\sigma_i^1,\sigma_i^1) = \frac{1}{2}\lim_{t\to 0}\frac{d}{d\sigma_i^1}G_t'(\sigma_i^1,\sigma_i^1).$$

Although another regularization was used, the above definition is in agreement with that of Ref. 6, when $\phi = 0$. To the best of my knowledge, the above vector amplitude is a new result.

Note added in proof: While this article was being typed, I received Ref. 18 where nonorientable topologies are also discussed within a first quantized approach. I thank C. P. Burgess for sending me the manuscript.

APPENDIX A: COMPUTATION OF DETERMINANTS

One needs to compute

$$\prod_{\substack{n_1 \text{ even}\\ n_2 \text{ even}}}{}' \left[\pi^2 n_1{}^2 + \frac{\pi^2 n_2^2}{\tau^2}\right]^{1/2} \qquad (A1)$$

and

$$\prod_{\substack{n_1 \text{ odd}\\ n_2 \text{ odd}}} \left[\pi^2 n_1{}^2 + \frac{\pi^2 n_2^2}{\tau^2}\right]^{1/2}. \qquad (A2)$$

The product in Eq. (A1) is readily written as

$$\prod{}' \left[(2\pi)^2 n_1{}^2 + \frac{(2\pi)^2 n_2^2}{\tau^2}\right]^{1/2}.$$

This has been computed in Ref. 7. The result is

$$\prod_{\substack{n_1 \text{ even}\\ n_2 \text{ even}}}{}' \left[\pi^2 n_1{}^2 + \frac{\pi^2 n_2^2}{\tau^2}\right]^{1/2} = \tau e^{-\pi\tau/6}\left[\prod_{n=1}^{\infty}(1-e^{-2\pi n\tau})\right]^2. \qquad (A3)$$

For the product in Eq. (A2) I use methods similar to those of Ref. 7. I consider

$$\frac{1}{2}\sum_{\substack{n_1,n_2\\ \text{both odd}}} \ln\left[\pi^2 n_1{}^2 + \frac{\pi^2 n_2^2}{\tau^2}\right] = -\frac{1}{2}\lim_{s\to 0}\frac{d}{ds}\sum_{n_1,n_2 \text{ both odd}}\left[\pi^2 n_1{}^2 + \frac{\pi^2 n_2^2}{\tau^2}\right]^{-s} = -\frac{1}{2}\lim_{s\to 0}\frac{d}{ds}\sum_{n_1 \text{ odd}}\left(\frac{\pi^2}{\tau^2}\right)^{-s}\sum_{n_2 \text{ odd}}[n_2{}^2 + n_1^2\tau^2]^{-s}. \qquad (A4)$$

The sum over n_2 is performed by complex integration:

$$\left(\frac{\pi^2}{\tau^2}\right)^{-s}\sum_{n_2 \text{ odd}}[n_2{}^2+n_1^2\tau^2]^{-s} = \left(\frac{\pi^2}{\tau^2}\right)^{-s}\int_{C^+ + C^-} dz\,\frac{i}{4}\tan\left(\frac{\pi z}{2}\right)[z^2+\tau^2 n_1^2]^{-s}$$
$$= \left(\frac{\pi^2}{\tau^2}\right)^{-s}\int_{C^+}\frac{dz}{4}\left(\frac{2}{1+e^{-i\pi z}}-1\right)[z^2+\tau^2 n_1^2]^{-s} + \left(\frac{\pi^2}{\tau^2}\right)^{-s}\int_{C^-}\frac{dz}{4}\left(\frac{-2}{1+e^{i\pi z}}+1\right)[z^2+\tau^2 n_1^2]^{-s}, \qquad (A5)$$

where C^+ (C^-) is a contour just above (below) the real axis from $+\infty$ to $-\infty$ ($-\infty$ to $+\infty$). One obtains

$$\left(\frac{\pi^2}{\tau^2}\right)^{-s}\sum_{n_2 \text{ odd}}[n_2{}^2+n_1^2\tau^2]^{-s} = (e^{i\pi s}-e^{-i\pi s})\left(\frac{\pi^2}{\tau^2}\right)^{-s}\times\int_0^{\infty}\frac{i\,dy}{1+e^{\pi\tau|n_1|}e^{\pi y}}\frac{1}{y^s(y+2\tau|n_1|)^s} - (e^{i\pi s}-e^{-i\pi s})\left(\frac{\pi^2}{\tau^2}\right)^{-s}\int_0^{\infty}\frac{i\,dy}{2}\frac{1}{y^s(y+2\tau|n_1|)^s}. \qquad (A6)$$

The second integral in the above equation (convergent only for $s > \frac{1}{2}$) is interpreted as

$$\frac{i}{2}(2\tau|n_1|)^{1-2s}\frac{\Gamma(1-s)\Gamma(2s-1)}{\Gamma(s)} = \frac{i}{2}(\tau|n_1|)^{1-2s}\frac{\Gamma(1-s)}{2s-1}\frac{\Gamma(s+\frac{1}{2})}{\sqrt{\pi}},$$

which now has a well-defined value at $s = 0$. One then obtains

$$\sum_{n_2 \text{ odd}}\ln[\tau^2 n_1{}^2+n_2^2] = 2\ln(1+e^{-\pi\tau|n_1|}) + \pi\tau\lim_{s\to 0}|n_1|^{1-2s},$$

from which it follows that

$$\ln\prod_{\substack{n_1 \text{ odd}\\ n_2 \text{ odd}}}\left[\pi^2 n_1{}^2 + \frac{\pi^2 n_2^2}{\tau^2}\right]^{1/2} = \frac{\pi\tau}{12} + 2\sum_{\substack{n>0\\ \text{odd}}}\ln(1+e^{-\pi n\tau}). \qquad (A7)$$

In the above, I have used

$$\sum_{n=1}^{\infty} n^{-s} = \zeta(s).$$

Equations (A3) and (A7) are Eqs. (27) and (28) in the main text.

APPENDIX B: EVALUATION OF THE PROPAGATOR

I will consider the Neumann problem

$$-\partial^a \partial_a G(\sigma,\sigma') = \delta(\sigma - \sigma') \tag{B1}$$

for the Möbius strip $0 \leqslant \sigma^a \leqslant 1$, $a = 1,2$. The propagator is defined by means of heat kernel regularization, i.e.,

$$G(\sigma,\sigma') = \lim_{z\to 0} \frac{1}{\Gamma(1+z)} \int_0^\infty dt\, t^z G_t(\sigma,\sigma'), \tag{B2}$$

where

$$G_t(\sigma,\sigma') = \sum_{n_1,n_2\ \text{even}} e^{i\pi n_1(\sigma^1 - \sigma'^1)} \cos(n_2\pi\sigma^2) \times \cos(n_2\pi\sigma'^2) e^{-t\pi^2 n_1^2} e^{-t(\pi^2 n_2^2/\tau^2)} + \sum_{n_1,n_2\ \text{odd}} e^{i\pi n_1(\sigma^1 - \sigma'^1)} \cos(n_2\pi\sigma^2) \times \cos(n_2\pi\sigma'^2) e^{-t\pi^2 n_1^2} e^{-t(\pi^2 n_2^2/\tau^2)}. \tag{B3}$$

Using the following straightforward application of Poisson summation formula,

$$\sum_{n\ \text{even}} e^{i\pi n x} e^{-t(\pi^2 n^2/\tau^2)} = \frac{\tau}{2\sqrt{\pi t}} \sum_n e^{-(\tau^2/4t)(n-x)^2}, \tag{B4a}$$

$$\sum_{n\ \text{odd}} e^{i\pi n x} e^{-t(\pi^2 n^2/\tau^2)} = \frac{\tau}{2\sqrt{\pi t}} \sum_n (-)^n e^{-(\tau^2/4t)(n-x)^2}, \tag{B4b}$$

one obtains for G_t,

$$G_t(\sigma,\sigma') = \frac{\tau}{4\pi t}\left(\sum_{\substack{n_1,n_2\\ \text{even}}} + \sum_{\substack{n_1,n_2\\ \text{odd}}}\right) \times e^{-(1/4t)(n_1 - (\sigma^1 - \sigma'^1))^2} e^{-(\tau^2/4t)(n_2 - (\sigma^2 - \sigma'^2))^2} + \frac{\tau}{4\pi t}\left(\sum_{\substack{n_1,n_2\\ \text{even}}} + \sum_{\substack{n_1,n_2\\ \text{odd}}}\right) \times e^{-(1/4t)(n_1 - (\sigma^1 - \sigma'^1))^2} e^{-(\tau^2/4t)(n_2 - (\sigma^2 + \sigma'^2))^2}. \tag{B5}$$

After integration, the propagator becomes

$$G(\sigma,\sigma') = \lim_{z\to 0} -\frac{\tau}{4\pi z}\frac{\Gamma(1-z)}{\Gamma(1+z)}\left(\sum_{\substack{n_1,n_2\\ \text{even}}} + \sum_{\substack{n_1,n_2\\ \text{odd}}}\right)\left[\frac{(n_1 - (\sigma^1 - \sigma'^1))^2 + \tau^2(n_2 - (\sigma^2 - \sigma'^2))^2}{4}\right]^z + \lim_{z\to 0} -\frac{\tau}{4\pi z}\frac{\Gamma(1-z)}{\Gamma(1+z)}\left(\sum_{\substack{n_1,n_2\\ \text{even}}} + \sum_{\substack{n_1,n_2\\ \text{odd}}}\right)\left[\frac{(n_1 - (\sigma^1 - \sigma'^1))^2 + \tau^2(n_2 - (\sigma^2 + \sigma'^2))^2}{4}\right]^z. \tag{B6}$$

Using methods similar to those described in Appendix A, one can easily obtain the formulas

$$\sum_{n\ \text{odd}} \left[\frac{(n-a)^2 + b^2}{4}\right]^z = 2z \ln|1 + e^{i\pi a} e^{-\pi b}| + z\pi b + o(z^2), \tag{B7}$$

$$\sum_{n\ \text{even}} \left[\frac{(n-a)^2 + b^2}{4}\right]^z = 2z \ln|1 - e^{i\pi a} e^{-\pi b}| + z\pi b + o(z^2). \tag{B8}$$

The propagator then becomes

$$G(\sigma,\sigma') = -\frac{\tau}{2\pi}\sum_n \ln|1 - (-)^n e^{i\pi(\sigma^1 - \sigma'^1)} e^{-\pi\tau|n - (\sigma^2 - \sigma'^2)|}||1 - (-)^n e^{i\pi(\sigma^1 - \sigma'^1)} e^{-\pi\tau|n - (\sigma^2 + \sigma'^2)|}| - \frac{\tau^2}{4}\sum_n |n - (\sigma^2 - \sigma'^2)| - \frac{\tau^2}{4}\sum_n |n - (\sigma^2 + \sigma'^2)|. \tag{B9}$$

In order to regularize the last two sums, I use

$$\sum_n \left|\frac{n-x}{2}\right| = \lim_{s\to 0}\sum_n \left[\frac{(n-x)^2}{4}\right]^{s+1/2} = \lim_{s\to 0} \frac{\sqrt{4\pi}\,\Gamma(1+s)}{\Gamma(-s-\frac{1}{2})} \sum_n \frac{e^{i2\pi n x}}{[(2\pi)^2 n^2]^{1+s}}, \tag{B10}$$

where a prescription has to be given to the zero mode in the sum of the right-hand side. The above equality is easily derived by considering the right-hand side as an unregularized one-dimensional propagator. Then its kernel regularization with the use of Poisson's summation formula yields the left-hand side of (B10).

When $x = 0$, the above equation requires that

$$\sum_n \frac{1}{(2\pi)^2 n^2} = -\zeta(-1) = \frac{1}{2\pi^2}\zeta(2). \tag{B11}$$

Clearly, this means that the zero mode must be removed from the sum. It is easy to establish by contour integration that

$$\sum_{n\neq 0} \frac{e^{2\pi n x}}{n^2} = \frac{\pi^2}{3} - 2\pi^2|x| + 2\pi^2 x^2. \quad \text{(B12)}$$

In the right-hand side of the above equation $|x| \leqslant 1$. This is always possible due to the original periodicity of the left-hand side. Once this result is substituted into Eq. (B9), Eq. (37a) in the main text follows.

Suppose that in Eq. (B6) $\sigma^1 - \sigma'^1 = \sigma^2 - \sigma'^2 = 0$. Then the term $n_2 = 0$ must be considered separately:

$$\lim_{z\to 0} -\frac{\tau}{4\pi z}\frac{\Gamma(1-z)}{\Gamma(1+z)} \sum_{n_1 \text{ even}} \left[\frac{n_1^{\,2}}{4}\right]^z$$

$$= \lim_{z\to 0} -\frac{\tau}{2\pi z}\frac{\Gamma(1-z)}{\Gamma(1+z)}\zeta(-2z)$$

$$= \lim_{z\to 0}\frac{\tau}{4\pi z} + \frac{\tau}{2\pi}\ln\left(\frac{e^\gamma}{2\pi}\right).$$

This is the regularization used in Eq. (37) of the main text.

The following final remark is in order. The combined use of heat kernel regularization, Poisson summation formula and zeta function regularization amounts to the removal of the zero mode from the propagator, as it was already noticed in the discussion following Eq. (B10). Indeed it can be shown that the propagator obtained above is equal to

$$\left(\sum_{\substack{n_1,n_2\\ \text{even}}}{}' + \sum_{\substack{n_1,n_2\\ \text{odd}}}\right) \frac{e^{i\pi n_1(\sigma^1 - \sigma'^1)}\cos(n_2\pi\sigma^2)\cos(n_2\pi\sigma'^2)}{\pi^2 n_1^{\,2} + \pi^2 n_2^{\,2}/\tau^2}.$$

This is not the propagator in the sense that it does not satisfy $-\partial^2 G = \delta$ but $-\partial^2 G = \delta - 1$. This means, in particular, that the normal value at the boundary is not the one obtained from Gauss' law. However, since one is actually interested in solving $-\partial^2 x = j$, and in our case $\int d^2\sigma j = 0$ (by momentum conservation), one is justified in using it. It may then be asked why one needs to use heat kernel regularization. The reason is that a regularized value of the propagator at coincident points is required.

[1]M. Kaku and K. Kikkawa, Phys. Rev. D **10**, 1110,1873 (1974). There is already a large recent literature on the subject. Some of the initial works are W. Siegel, Phys. Lett. B **142**, 276 (1984); **149**, 157,162 (1984); D. Friedan, UC-EFI preprint 85-27, 1985; T. Banks and M. Peskin, in *Anomalies, Topology and Geometry*, edited by A. White (World Scientific, Singapore, 1985).

[2]C. S. Hue, B. Sakita, and M. A. Virasoro, Phys. Rev. D **2**, 2857 (1970).

[3]A. M. Polyakov, Phys. Lett. B **103**, 207 (1981).

[4]Y. Goldschmidt and C.-I. Tan, Phys. Lett. B **114**, 257 (1982); D. Friedan, in *Recent Advances in Field Theory and Statistical Mechanics*, Les Houches, 1982, edited by J.-B. Zuber and R. Stora (North-Holland, Amsterdam, 1984).

[5]O. Alvarez, Nucl. Phys. B **216**, 125 (1983).

[6]H. Aoyama, A. Dhar, and M. A. Namazie, Nucl. Phys. B **267**, 605 (1986); D. Friedan and S. Shenker, UC-EFI preprint, 1985.

[7]J. Polchinski, Commun. Math. Phys. **104**, 37 (1986).

[8]G. Moore and P. Nelson, Nucl. Phys. B **266**, 58 (1986).

[9]E. D'Hoker and D. H. Pong, Nucl. Phys. B **269**, 205 (1986).

[10]G. Moore, P. Nelson, and J. Polchinski, Phys. Lett. B **169**, 47 (1986); P. Nelson and G. Moore, Nucl. Phys. B **274**, 509 (1986).

[11]A. Cohen, G. Moore, P. Nelson, and J. Polchinski, Nucl. Phys. B **267**, 143 (1986).

[12]E. S. Fradkin and A. A. Tseytlin, Phys. Lett. B **158**, 316 (1985); **160**, 69 (1985); JETP Lett. **41**, 206 (1985).

[13]B. Fridling and A. Jevicki, Phys. Lett. B **174**, 75 (1986).

[14]J. P. Rodrigues, Phys. Lett. B **178**, 350 (1986).

[15]M. Green and J. H. Schwarz, Phys. Lett. B **151**, 21 (1985).

[16]See for instance the discussion contained in J. H. Schwarz, Phys. Rep. **89**, 223 (1982).

[17]J. P. Rodrigues (in preparation).

[18]C. P. Burgess and T. R. Morris, Nucl. Phys. B **291**, 256 (1987).

Pole term decomposition of the resolvent kernel in Fredholm's form

Yoshiko Matsui
Department of Applied Physics, Tokyo University of Agriculture and Technology, Naka-machi, Koganei-shi, Tokyo, Japan

(Received 13 February 1987; accepted for publication 24 June 1987)

The resolvent kernel represented by a ratio of two Fredholm's series can be reformed into a sum of two ratios, only one of which contains Fredholm's determinant in its denominator. The general results presented are then used to apply Faddeev's residue prescription in the case of the rank N separable potential for determining those three-body scattering amplitudes in which a two-particle bound state is in the initial and/or final configuration. As a numerical example, the triton binding energy and the doublet scattering length for the Tabakin potential are calculated.

I. INTRODUCTION

Faddeev[1] established the scattering theory for the three-body system in the framework of the spectral theory. Practical applications to various fields of the nuclear reaction[2,3] have followed. Although the nonlocal separable potential has played an important role, the multichannel scattering problem related to the general separable potential with infinite or finite rank is not solved. We say that the infinite[4,5] or finite rank separable potential is not general in the case when its first form factor is determined from the bounded state wave function and the others are chosen to be orthogonal to the first.[6] The Tabakin potential[7] or Mongan potential[8] are examples of the general separable potential. With only rank 2 these potentials may well reproduce the realistic interaction, and yet the application of them to the practical calculation for three-body scattering has not received much attention. To consider the question, a more pertinent form of the resolvent kernel represented in Fredholm's series is desirable: from such a form we can derive Faddeev's equation for the general finite (or infinite) separable potential.

As was emphasized by Osborn[9] and many other authors,[10–12] Faddeev developed his theory by using the residues instead of the three-body scattering amplitudes themselves, because the amplitudes contain the primary singularities, and he then could define the elastic, rearrangement, or breakup amplitude of the three-body reaction. These primary singularities arise from the bound-state pole terms of the two-body T matrices when they are expressed in the spectral expansion.

To proceed with the practical theory for the three-body scattering with the general separable potential, a pole decomposition of the resolvent kernel in the Fredholm's form is needed, by analogy with Faddeev's residue prescription. In this paper, the resolvent kernel represented by a ratio with the Fredholm's determinant for its denominator is transformed into two ratios: only one of them contains the Fredholm's determinant in its denominator which causes the bound-state pole.

In Sec. II, the bound-state pole decomposition of the resolvent kernel in Fredholm's series is derived, with proofs given in the Appendices. This decomposition of the two-body T matrix for the Tabakin potential is applied in Sec. III to derive the Faddeev equation for the three-nucleon system. A numerical example is given.

II. POLE TERM DECOMPOSITION

The first partial wave off-shell T matrix satisfies the off-shell Lippmann–Schwinger (LS) equation

$$T_l(p,p';z) = V_l(p,p') + 4\pi \int_0^\infty V_l(p,p'')\left[z-\frac{p''^2}{2m}\right]^{-1} \times T_l(p'',p';z)p''^2\,dp''. \tag{1}$$

For a fixed p' the LS equation can be studied as a Fredholm's integral equation of the second kind

$$\varphi(x) = f(x) + \lambda \int_\Omega N(x,y)\varphi(y)dy. \tag{2}$$

Here we take the set Ω as a bounded interval instead of the interval $[0,\infty)$ to avoid some considerations of the convergence, since the separable potential used in our study is proved to make the kernel $N(x,y)$ a compact operator.[13]

Hereafter we proceed along Pogorzelski's[14] argument, with his notation. Fredholm's first theorem is described as follows: Fredholm's equation (2) of the second kind, under the assumption that the function $f(x)$ and $N(x,y)$ are integrable, has in the case $D(\lambda)\neq 0$ a unique solution, which is of the form

$$\varphi(x) = f(x) + \lambda \int_\Omega R(x,y,\lambda) f(y)dy, \tag{3}$$

where the resolvent kernel R is a meromorphic function of the parameter λ, being the ratio of two entire functions of the parameter λ,

$$R(x,y,\lambda) = D\begin{pmatrix} x \\ y \end{pmatrix}; \lambda\Big)[D(\lambda)]^{-1}, \tag{4}$$

defined by Fredholm's series of the form

$$D(\lambda) = 1 + \sum_{p=1}^{\infty} \frac{(-\lambda)^p}{p!}$$

 0022-2488/87/112677-06$02.50

$$\times \int_\Omega \cdots \int_\Omega N\begin{pmatrix} s_1,s_2,\ldots,s_p \\ s_1,s_2,\ldots,s_p \end{pmatrix} ds_1\, ds_2 \cdots ds_p, \qquad (5)$$

$$D\begin{pmatrix} x \\ y \end{pmatrix};\lambda) = N(x,y) + \sum_{p=1}^{\infty} \frac{(-\lambda)^p}{p!} \times \int_\Omega \cdots \int_\Omega N\begin{pmatrix} x,s_1,\ldots,s_p \\ y,s_1,\ldots,s_p \end{pmatrix} ds_1 \cdots ds_p. \qquad (6)$$

These series converge for all values of λ. The integrands in Eqs. (5) and (6) are defined as follows:

$$N\begin{pmatrix} s_1,s_2,\ldots,s_p \\ s_1,s_2,\ldots,s_p \end{pmatrix} = \begin{vmatrix} N(s_1,s_1) & N(s_1,s_2) & \cdots & N(s_1,s_p) \\ N(s_2,s_1) & N(s_2,s_2) & \cdots & N(s_2,s_p) \\ \vdots & \vdots & \vdots & \vdots \\ N(s_p,s_1) & N(s_p,s_2) & \cdots & N(s_p,s_p) \end{vmatrix}, \qquad (7)$$

$$N\begin{pmatrix} x,s_1,s_2,\ldots,s_p \\ y,s_1,s_2,\ldots,s_p \end{pmatrix} = \begin{vmatrix} N(x,y) & N(x,s_1) & N(x,s_2) & \cdots & N(x,s_p) \\ N(s_1,y) & N(s_1,s_1) & N(s_1,s_2) & \cdots & N(s_1,s_p) \\ N(s_2,y) & N(s_2,s_1) & N(s_2,s_2) & \cdots & N(s_2,s_p) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ N(s_p,y) & N(s_p,s_1) & N(s_p,s_2) & \cdots & N(s_p,s_p) \end{vmatrix}. \qquad (8)$$

We state our result for the resolvent kernel $R(x,y,\lambda)$.

Theorem 1: Under the condition

$$D(\lambda) \neq 0, \quad D\begin{pmatrix} s \\ t \end{pmatrix};\lambda) \neq 0 \quad \text{for } s,t \in \Omega,$$

we have the following relation:

$$\frac{D\begin{pmatrix} x \\ y \end{pmatrix};\lambda)}{D(\lambda)} = \frac{D\begin{pmatrix} x \\ t \end{pmatrix};\lambda) D\begin{pmatrix} s \\ y \end{pmatrix};\lambda)}{D(\lambda) D\begin{pmatrix} s \\ t \end{pmatrix};\lambda)} + \frac{D\begin{pmatrix} x,s \\ y,t \end{pmatrix};\lambda)}{D\begin{pmatrix} s \\ t \end{pmatrix};\lambda)}. \qquad (9)$$

Proof: See Appendix A.

Before stating our Theorem 2, we must enunciate Fredholm's second theorem: If λ_0 is an eigenvalue of rank r, then the homogenous integral equation

$$\varphi(x) = \lambda_0 \int_\Omega N(x,y)\varphi(y)dy \qquad (10)$$

possesses r independent solutions

$$\varphi_i(x) = \frac{D\begin{pmatrix} s_1,s_2,\ldots,s_{i-1},x,s_{i+1},\ldots,s_r \\ t_1,t_2,\ldots,t_{i-1},t_i,t_{i+1},\ldots,t_r \end{pmatrix};\lambda)}{D\begin{pmatrix} s_1,\ldots,s_r \\ t_1,\ldots,t_r \end{pmatrix};\lambda)} \quad (i = 1,\ldots,r), \qquad (11)$$

for $s_1,\ldots,s_r$, $t_1,\ldots,t_r$ satisfying

$$D\begin{pmatrix} s_1,\ldots,s_r \\ t_1,\ldots,t_r \end{pmatrix};\lambda) \neq 0.$$

Simultaneously we may refer to the homogeneous associated equation

$$\psi(x) = \int_\Omega N(y,x)\psi(y)dy, \qquad (12)$$

which has r characteristic solutions

$$\psi_i(x) = \frac{D\begin{pmatrix} s_1,\ldots,s_{i-1},s_i,s_{i+1},\ldots,s_r \\ t_1,\ldots,t_{i-1},y,t_{i+1},\ldots,t_r \end{pmatrix};\lambda)}{D\begin{pmatrix} s_1,\ldots,s_r \\ t_1,\ldots,t_r \end{pmatrix};\lambda)} \quad (i = 1,\ldots,r), \qquad (13)$$

for $s_1,\ldots,s_r$, $t_1,\ldots,t_r$ satisfying

$$D\begin{pmatrix} s_1,\ldots,s_r \\ t_1,\ldots,t_r \end{pmatrix};\lambda) \neq 0.$$

With regard to the first term of right-hand side of Eq. (9) in Theorem 1 we have the following result.

Theorem 2: When λ_0 is an eigenvalue of rank 1 and then the resolvent $R(x,y,\lambda_0)$ has a pole of order 1 at λ_0, we designate the residue of R at this pole $-c(x,y)$. Then it has the form

$$c(x,y) = \frac{D\begin{pmatrix} x \\ t \end{pmatrix};\lambda_0) D\begin{pmatrix} s \\ y \end{pmatrix};\lambda_0)}{D^{(1)}(\lambda_0) D\begin{pmatrix} s \\ t \end{pmatrix};\lambda_0)} \quad \left[D\begin{pmatrix} s \\ t \end{pmatrix};\lambda_0) \neq 0\right], \qquad (14)$$

satisfying the relation

$$\int_\Omega \frac{D\begin{pmatrix} x \\ t \end{pmatrix};\lambda_0) D\begin{pmatrix} s \\ x \end{pmatrix};\lambda_0)}{D^{(1)}(\lambda_0) D\begin{pmatrix} s \\ t \end{pmatrix};\lambda_0)} dx = -1, \qquad (15)$$

where

$$D^{(1)}(\lambda) = \frac{d}{d\lambda} D(\lambda).$$

Proof: See Appendix B.

When the eigenvalue λ_0 equals 1, this theorem implies that the first term of the right-hand side of Eq. (9) corresponds to the pole term of the two-body T matrix in spectral representation.

To see the meaning of the final term of Eq. (9), we need Fredholm's third theorem: When λ is an eigenvalue of rank r, a necessary and sufficient condition for the existence of a solution of Eq. (2) is the orthogonality of the given function $f(x)$ to r characteristic solutions ψ_i of the associated homogeneous equation (12) corresponding to λ. Further, the theorem proceeds to state: If this condition is satisfied, then the general solution of Eq. (2) has the form

$$\varphi(x) = f(x) + \lambda \int_\Omega U(x,y,\lambda) f(y) dy + \sum_{j=1}^{r} c_j \varphi_j(x), \tag{16}$$

where the c_j are constants and the $\varphi_j(x)$ are the characteristic solutions (11) of the homogeneous equation (10) and $U(x,y)$ has the form

$$U(x,y,\lambda) = \frac{D\begin{pmatrix} x,s_1,...,s_r \\ y,t_1,...,t_r \end{pmatrix}; \lambda)}{D\begin{pmatrix} s_1,...,s_r \\ t_1,...,t_r \end{pmatrix}; \lambda)}. \tag{17}$$

From this theorem we see that the final term of our representation (9) for the resolvent is equal to the extended $U(x,y,\lambda)$ when λ is not the eigenvalue.

III. APPLICATION TO THE THREE-BODY PROBLEM

In this section we study the utility of the pole term decomposition acquired in the previous section by applying it for the calculation of the three-body problem.

In the LS equation (1) the general form of an s-wave spin-dependent potential is assumed to be expressed as a finite series of nonlocal separable potentials,

$$V_\nu(p,p') = \sum_{k=1}^{n} v_{\nu k}(p) v_{\nu k}(p'), \tag{18}$$

where the label ν specifies an antisymmetric spin–isospin state of the two intermediately coupled nucleons of a three-nucleon system: it denotes that the pair has the spin–isospin, $st = 10(01)$ for $\nu = 0(1)$, respectively. As we consider the s-wave potential, the l index will be omitted in the LS equation.

From the results of the previous section together with the Appendices, if we represent the total spin and isospin of three-nucleon system as σ and τ, the analytical solution of the LS equation is given by the equation

$$T(p,p';z) = \sum_{\nu=0}^{1} \sum_m |\sigma\tau\nu\rangle g_{\nu m}(p;z) f_{\nu m}(p';z) \langle\sigma\tau\nu|, \tag{19}$$

where

$$g_{\nu 1}(p;z) = D_n\begin{pmatrix} p \\ s \end{pmatrix};1\Big),$$

$$f_{\nu 1}(p;z) = D_n\begin{pmatrix} s \\ p \end{pmatrix};1\Big)[d_{\nu 1}(z)]^{-1},$$

$$d_{\nu 1}(z) = D_n(1) D_n\begin{pmatrix} s \\ s \end{pmatrix};1\Big),$$

$$g_{\nu 2}(p;z) = \{v_{\nu 1}(p) v_{\nu 2}(s) - v_{\nu 2}(p) v_{\nu 1}(s)\} \times \Big[D_n\begin{pmatrix} s \\ s \end{pmatrix};1\Big)\Big]^{-1},$$

$$f_{\nu 2}(p;z) = \begin{vmatrix} v_{\nu 1}(p) & v_{\nu 1}(s) & -a_{13} & \cdots & -a_{1n} \\ v_{\nu 2}(p) & v_{\nu 2}(s) & -a_{23} & \cdots & -a_{2n} \\ v_{\nu 3}(p) & v_{\nu 3}(s) & 1-a_{33} & \cdots & -a_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ v_{\nu n}(p) & v_{\nu n}(s) & -a_{n3} & \cdots & 1-a_{nn} \end{vmatrix}, \tag{20}$$

and so on. The index m of the second sum of Eq. (19) runs on the combination such that we take two numbers of the set $1,2,...,n$. Here $|\sigma\tau\nu\rangle$ is the unified spin–isospin function of the three-nucleon system. In the determinant of Eq. (20), a_{ij} is defined as follows:

$$a_{ij} = 4\pi \int_0^\infty \frac{v_{\nu i}(p) v_{\nu j}(p)}{z - p^2/2m} p^2\, dp.$$

The first partial wave three-nucleon off-shell scattering amplitude satisfies the Faddeev equation

$$w^{\sigma\tau;l}_{\nu' m';\nu m}(p',p;z) = U^{\sigma\tau;l}_{\nu' m';\nu m}(p',p;z) + \sum_{\nu'' m''} \int_0^\infty U^{\sigma\tau;l}_{\nu' m';\nu'' m''}(p',p'';z) \times w^{\sigma\tau;l}_{\nu'' m'';\nu m}(p'',p;z) p''^2\, dp'', \tag{21}$$

with the potential matrix

$$U^{\sigma\tau;l}_{\nu' m';\nu m}(p',p;z) = 2(2/\sqrt{3})^3 \langle\sigma\tau\nu'|\sigma\tau\nu\rangle \times \int_{-1}^{1} \frac{f_{\nu' m'}(p_1,t') g_{\nu m}(p_2,t)}{z - (1/2m)(p^2 + p_1^2)} P_l(x) dx, \tag{22}$$

where

$$p_1^2 = \tfrac{1}{3}p^2 + \tfrac{4}{3}(p'^2 + pp'x), \quad p_2^2 = \tfrac{1}{3}p'^2 + \tfrac{4}{3}(p^2 + pp'x),$$

$$t = z - p^2/2m, \quad t' = z - p'^2/2m.$$

Here $w^{\sigma(1/2);l}_{01;01}$ is related to the completely antisymmetrized off-shell amplitude $\widetilde{w}^{\sigma(1/2);l}_{01;01}$ of elastic n-d scattering with given σ by the following form:

$$w^{\sigma(1/2);l}_{01;01}(p',p;z) = \widetilde{w}^{\sigma(1/2);l}_{01;01}(p',p;z)/d_{01}(t'). \tag{23}$$

It follows, as in Osborn,[9] that the equation for the off-shell elastic amplitude $\widetilde{w}^{\sigma(1/2);l}_{01;01}$ can be obtained by inserting relation (23) and the second equation of Eq. (20) into Eq. (21),

$$\widetilde{w}^{\sigma(1/2);l}_{01;01}(p',p;z) = \widetilde{U}^{\sigma(1/2);l}_{01;01}(p',p;z) + \int_0^\infty \widetilde{U}^{\sigma(1/2);l}_{01;01}(p',p'';z) \times d^{-1}_{01}(t'')\widetilde{w}^{\sigma(1/2);l}_{01;01}(p'',p;z)\, p''^2\, dp'' + \sum_{\substack{\nu m \\ (\nu m \neq 01)}} \int_0^\infty \widetilde{U}^{\sigma(1/2);l}_{01;\nu m}(p',p'';z) \times w^{\sigma(1/2);l}_{\nu m;01}(p'',p;z) p''^2\, dp'', \tag{24}$$

with the potential matrix

$$\widetilde{U}^{\sigma(1/2);l}_{01;\nu m}(p',p;z) = 2\Big(\frac{2}{\sqrt{3}}\Big)^3 \Big\langle \sigma\frac{1}{2}0\Big|\sigma\frac{1}{2}\nu\Big\rangle$$

$$\times\int_{-1}^{1}\frac{g_{01}(p_1,t')g_{vm}(p_2,t')}{z-(1/2m)(p^2+p_1^2)}P_l(x)dx.$$

We obtain a closed set of equations once we add to Eq. (24) a linear equation giving $w^{\sigma(1/2);l}$ $(\mu m\neq 0,1)$ in terms $w^{\sigma(1/2);l}$ and $\tilde{w}^{\sigma(1/2);l}$. The necessary equation is driven from Eq. (21),

$$\begin{aligned}
&w^{\sigma(1/2);l}_{vm;01}(p',p;z)\\
&\quad=U^{\sigma(1/2);l}_{vm;01}(p',p;z)+\int_0^\infty U^{\sigma(1/2);l}_{vm;01}(p',p'';z)\\
&\qquad\times\tilde{w}^{\sigma(1/2);l}_{01;01}(p'',p;z)\,p''^2\,dp''\\
&\qquad+\sum_{\substack{v'm'\\(v'm'\neq 01)}}\int_0^\infty U^{\sigma(1/2);l}_{vm;v'm'}(p',p'';z)\\
&\qquad\times w^{\sigma(1/2);l}_{v'm';01}(p'',p;z)\,p''^2\,dp''.
\end{aligned}\tag{25}$$

Now Eqs. (24) and (25) are Osborn's set of solvable coupled integral equations which are expressed by Fredholm's series. The breakup amplitude is constructed from these amplitudes $\tilde{w}$ and w. Forthwith we can embark on our numerical analysis of this set of equations.

Using the Tabakin potential,[7] the triton binding energy and the doublet scattering length are calculated. The results are -9.29 MeV and -0.1132 F, respectively. Figure 1 compares these with the experimental values and results of other calculations. In spite of omitting the tensor force, the Tabakin potential has an adequate agreement with the experimental value.

In conclusion, the application of our expression of the resolvent kernel to four-nucleon scattering can be anticipated.

ACKNOWLEDGMENTS

The author wishes to express her hearty thanks to Professor S. Fukutomi for useful discussions. Also, the author wishes to thank the staff of the computer room of the Institute for Nuclear Study for the use of the computer, Facom M380.

APPENDIX A: PROOF OF THEOREM 1

In this appendix, the proof of Theorem 1 is shown.

First we consider the case that the kernel of the Fredholm integral equation (2) has a degenerate kernel:

$$N(x,y)=\sum_{i=1}^{n}k_i(x)l_i(y).\tag{A1}$$

The first and second Fredholm's series of Eq. (2) with the degenerate kernel are defined by the formula

$$D_n(\lambda)=\begin{vmatrix}1-\lambda a_{11} & -\lambda a_{12} & \cdots & -\lambda a_{1n}\\ -\lambda a_{21} & 1-\lambda a_{22} & \cdots & -\lambda a_{2n}\\ \vdots & \vdots & \vdots & \vdots\\ -\lambda a_{n1} & -\lambda a_{n2} & \cdots & 1-\lambda a_{nn}\end{vmatrix}\tag{A2}$$

and

$$D_n\binom{x}{y};\lambda\Big)=-\begin{vmatrix}0 & k_1(x) & k_2(x) & \cdots & k_n(x)\\ l_1(y) & 1-\lambda a_{11} & -\lambda a_{12} & \cdots & -\lambda a_{1n}\\ l_2(y) & -\lambda a_{21} & 1-\lambda a_{22} & \cdots & -\lambda a_{2n}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ l_n(y) & -\lambda a_{n1} & -\lambda a_{n2} & \cdots & 1-\lambda a_{nn}\end{vmatrix}\tag{A3}$$

and

$$a_{ij}=\int_\Omega k_i(x)l_j(x)dx.$$

Defining $L_i(y)$ by

$$L_i(y)=(-)^{i+1}\begin{vmatrix}l_1(y) & 1-\lambda a_{11},\ldots,-\lambda a_{1,i-1},-\lambda a_{1,i+1},\ldots,-\lambda a_{1n}\\ l_2(y) & -\lambda a_{21},\ldots,-\lambda a_{2,i-1},-\lambda a_{2,i+1},\ldots,-\lambda a_{2n}\\ \vdots & \vdots\\ l_n(y) & -\lambda a_{n1},\ldots,-\lambda a_{n,i-1},-\lambda a_{n,i+1},\ldots,1-\lambda a_{nn}\end{vmatrix}\quad(i=1,2,\ldots,n),\tag{A4}$$

$D_n\binom{x}{y};\lambda)$ can be expanded with $L_i(y)$

$$D_n\Big(\begin{matrix}x\\y\end{matrix};\lambda\Big)=\sum_{i=1}^{n}k_i(x)L_i(y).\tag{A5}$$

From this we have

$$\begin{aligned}
&D_n\Big(\begin{matrix}x\\y\end{matrix};\lambda\Big)D_n\Big(\begin{matrix}s\\t\end{matrix};\lambda\Big)-D_n\Big(\begin{matrix}x\\t\end{matrix};\lambda\Big)D_n\Big(\begin{matrix}s\\y\end{matrix};\lambda\Big)\\
&\quad=\sum_{k=1}^{n}k_k(x)L_k(y)\sum_{l=1}^{n}k_l(s)L_l(t)\\
&\qquad-\sum_{k=1}^{n}k_k(x)L_k(t)\sum_{l=1}^{n}k_l(s)L_l(y)\\
&\quad=\sum_{k<l}\{k_k(x)k_l(s)-k_l(x)k_k(s)\}\\
&\qquad\times\{L_k(y)L_l(t)-L_l(y)L_k(t)\}.
\end{aligned}\tag{A6}$$

Let $\tilde{h}_{kl}$ be (kl)-cofactor in the Fredholm's determinant $D_n(\lambda)$. Then we can get an analogous expression

$$L_k(y)L_l(t)-L_l(y)L_k(t)$$

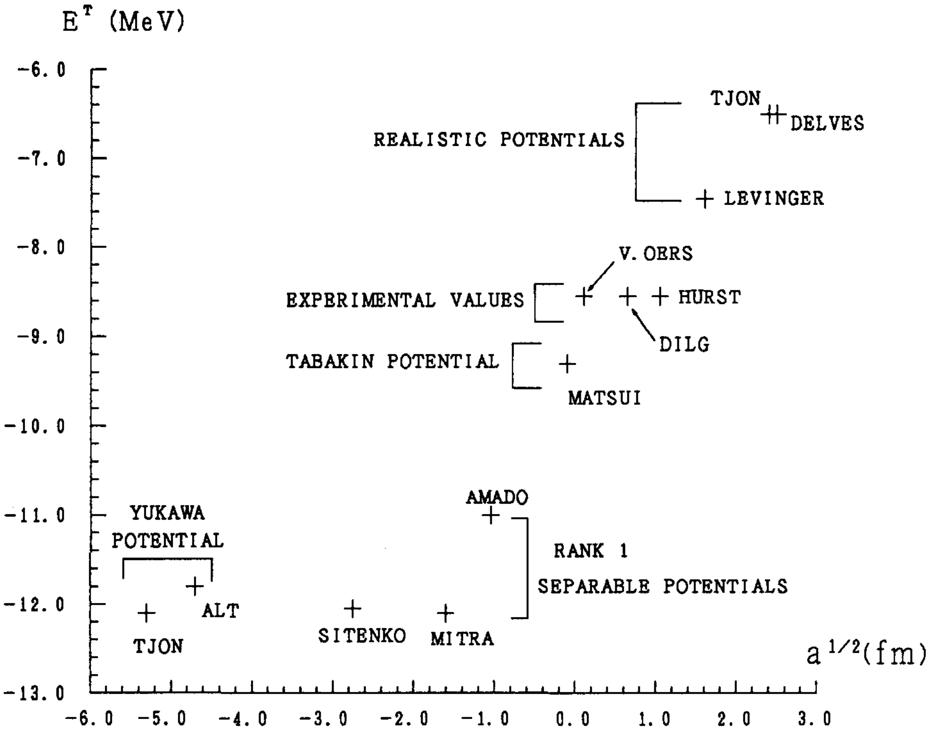

FIG. 1. Theoretical and experimental results of the triton binding energy plotted versus doublet scattering length. Adapted from Fig. 27 of Ref. 15.

$$= \sum_{i<j} \{\tilde{h}_{ik}\tilde{h}_{jl} - \tilde{h}_{il}\tilde{h}_{jk}\}\{l_i(y)l_j(t) - l_j(y)l_i(t)\}. \tag{A7}$$

In general, Jacob's theorem related to the minor of the determinant Δ is described as follows: Let $\Delta\binom{i}{k}$ be an $(n-1)\times(n-1)$ minor obtained by omitting the ith row and kth column of Δ and $\Delta\binom{i\,j}{k\,l}$ be an $(n-2)\times(n-2)$ minor taken off the ith and jth rows and kth and lth columns of Δ, to obtain the following relation:

$$\Delta\cdot\Delta\begin{pmatrix} i & j \\ k & l \end{pmatrix} = \Delta\binom{i}{k}\Delta\binom{j}{l} - \Delta\binom{i}{l}\Delta\binom{j}{k} \quad (i<j,k<l).$$

Using $D_n(\lambda)$ instead of Δ in this theorem, we can proceed with the derivation

$$\tilde{h}_{ik}\tilde{h}_{jl} - \tilde{h}_{il}\tilde{h}_{jk} = D_n(\lambda)\tilde{H}\begin{pmatrix} i & j \\ k & l \end{pmatrix}, \tag{A8}$$

where $\tilde{H}\binom{ij}{kl}$ is equal to $\Delta\binom{ij}{kl}$ multiplied by the factor $(-)^{i+j+k+l}$ in Jacob's theorem.

Inserting Eqs. (A7) and (A8) into (A6) and using Laplace's expansion of the determinant, the last form of the left-hand side of Eq. (A6) can be expressed as follows:

$$D_n\binom{x}{y};\lambda\Big)D_n\binom{s}{t};\lambda\Big) - D_n\binom{x}{t};\lambda\Big)D_n\binom{s}{y};\lambda\Big)$$

$$= D_n(\lambda)\sum_{i<j}\sum_{k<l}\begin{vmatrix} l_i(y) & l_i(t) \\ l_j(y) & l_j(t) \end{vmatrix} \times\begin{vmatrix} k_k(x) & k_l(x) \\ k_k(s) & k_l(s) \end{vmatrix}\tilde{H}\begin{pmatrix} i & j \\ k & l \end{pmatrix} = D_n(\lambda)D_n\begin{pmatrix} x,s \\ y,t \end{pmatrix};\lambda\Big).$$

Therefore we can derive Eq. (9) of Theorem 1 in the case of the finite rank kernel.

In the case of the general kernel, the proof of Eq. (9) is completed by using the theory of limiting process given in Ref. 14: Given a sequence of bounded continuous kernels $\{N_n(x,y)\}$ converging uniformly to the function $N(x,y)$,

$$N_n(x,y)\to N(x,y),$$

for x and y lying in the domain Ω, then the sequences of Fredholm's series D_n, etc., tend to D, etc.,

$$D_n(\lambda)\to D(\lambda),\quad D_n\begin{pmatrix} x \\ y \end{pmatrix};\lambda\Big)\to D\begin{pmatrix} x \\ y \end{pmatrix};\lambda\Big),$$

$$D_n\begin{pmatrix} x & s \\ y & t \end{pmatrix};\lambda\Big)\to D\begin{pmatrix} x & s \\ y & t \end{pmatrix};\lambda\Big),$$

under the assumption that λ is not an eigenvalue of any of the kernels $N_n(x,y)$ and $N(x,y)$.

APPENDIX B: PROOF OF THEOREM 2

In this appendix, the proof of Theorem 2 is given. As in Appendix A, we start with the degenerate kernel.

As is clear from its expression given by Eq. (9), the singular part of the resolvent $R_n(x,y,\lambda)$ is only the first term of the right-hand side of Eq. (9). If λ_0 is a zero of $D_n(\lambda)$ with multiplicity 1, the residue of the resolvent is equal to

$$\lim_{\lambda\to\lambda_0}(\lambda-\lambda_0)R_n(x,y,\lambda) = \frac{D_n\binom{x}{t};\lambda_0\Big)D_n\binom{s}{y};\lambda_0\Big)}{D_n^{(1)}(\lambda_0)D_n\binom{s}{t};\lambda_0\Big)}. \tag{B1}$$

Therefore Eq. (14) of Theorem 2 is proved.

Next consider the last part of the proof of Theorem 2. Let h_{ij} be a element of $D_n(\lambda_0)$ and $\tilde{h}_{ij}$, cofactor of h_{ij}. Then we have

$$\sum_{k=1}^{n} h_{ik}\tilde{h}_{jk} = \begin{cases} D_n(\lambda_0) & (i=j), \\ 0 & (i\neq j). \end{cases}$$

These formulas, together with the assumption of the theorem that λ_0 is a zero of $D_n(\lambda_0) = 0$ with multiplicity 1, give a unique solution $(\tilde{h}_{i1}, \tilde{h}_{i2}, ..., \tilde{h}_{in})^T$ for any $i = 1,...,n$ of the equation

$$H_n \mathbf{x} = 0,$$

where $H_n = (h_{ij})$ and $\mathbf{x}$ is an unknown vector. Then we have the following relations:

$$\frac{\tilde{h}_{i1}}{\tilde{h}_{j1}} = \frac{\tilde{h}_{i2}}{\tilde{h}_{j2}} = \cdots = \frac{\tilde{h}_{in}}{\tilde{h}_{jn}},$$

and (B2)

$$\frac{\tilde{h}_{1i}}{\tilde{h}_{1j}} = \frac{\tilde{h}_{2i}}{\tilde{h}_{2j}} = \cdots = \frac{\tilde{h}_{ni}}{\tilde{h}_{nj}},$$

for any $i, j = 1,2,...,n$.

Relation (B2) allows the following expansion of Fredholm's minor

$$D_n\binom{x}{t};\lambda_0) = \sum_{i,j=1}^{n} k_i(x)\tilde{h}_{1i} l_j(t) \frac{\tilde{h}_{ji}}{\tilde{h}_{1i}} = \sum_{i=1}^{n} k_i(x)\tilde{h}_{1i} L(t)$$

and similarly

$$D_n\left(\begin{matrix} s \\ y \end{matrix};\lambda_0\right) = \sum_{i=1}^{n} l_i(y)\tilde{h}_{1i} K(s),$$

$$D_n\left(\begin{matrix} s \\ t \end{matrix};\lambda_0\right) = \sum_{i=1}^{n} k_i(s)\tilde{h}_{1i} L(t), \qquad \text{(B3)}$$

where $L(t)$ or $K(s)$ are independent of index i.

For the derivation of Eq. (15) of Theorem 2 it is sufficient to show that

$$\int_\Omega \frac{D_n\left(\begin{matrix} x \\ t \end{matrix};\lambda_0\right) D_n\left(\begin{matrix} s \\ x \end{matrix};\lambda_0\right)}{D_n\left(\begin{matrix} s \\ t \end{matrix};\lambda_0\right)} dx = -D^{(1)}(\lambda_0). \qquad \text{(B4)}$$

Inserting expressions (B3) into each Fredholm's minor of the left-hand side of Eq. (B4), we can perform the integration of (B4):

$$\int_\Omega \frac{D_n\left(\begin{matrix} x \\ t \end{matrix};\lambda_0\right) D_n\left(\begin{matrix} s \\ x \end{matrix};\lambda_0\right)}{D_n\left(\begin{matrix} s \\ t \end{matrix};\lambda_0\right)} dx = \sum_{i,j=1}^{n} a_{ji}\tilde{h}_{ji}. \qquad \text{(B5)}$$

Let us now consider the right-hand side of Eq. (B4) and use the well-known relation given by Fredholm

$$\frac{d^k D_n(\lambda)}{d\lambda^k} = (-)^k \int_\Omega \cdots \int_\Omega D_n\left(\begin{matrix} x_1,...,x_k \\ x_1,...,x_k \end{matrix};\lambda\right) \times dx_1 \cdots dx_k \quad (1 \leqslant k \leqslant n) \qquad \text{(B6)}$$

between the kth derivative of Fredholm's determinant $D_n(\lambda)$ and Fredholm's minor of order k. Apply this relation in the case of $k = 1$ to get

$$-D^{(1)}(\lambda_0) = \int_\Omega D_n\left(\begin{matrix} x \\ x \end{matrix};\lambda_0\right) dx = \sum_{i,j=1}^{n} a_{ji}\tilde{h}_{ji}. \qquad \text{(B7)}$$

By comparing (B7) with (B5) we can show (B4).

We can also complete the proof in the general case along a procedure similar to that of Appendix A.

Hence Theorem 2 shows that the general property of the residue of the resolvent is also fulfilled by the one concretely expressed by Fredholm's series.

[1]L. D. Faddeev, *Mathematical Aspects of the Three-Body Problem in the Quantum Scattering Theory* (Davey, New York, 1965).

[2]C. Lovelace, Phys. Rev. **135**, B1225 (1964).

[3]E. O. Alt, P. Grassberger, and W. Sandhas, Nucl. Phys. B **2**, 167 (1967).

[4]C. Daskaloyannis, J. Math. Phys. **22**, 1861 (1981).

[5]Y. Matsui, Phys. Rev. C **26**, 2620 (1982); **29**, 253 (1984).

[6]P. Doleschall, Nucl. Phys. A **201**, 264 (1973); P. Doleschall, W. Grüebler, V. König, P. A. Schmelzbach, F. Sperisen, and B. Jenny, *ibid.* **380**, 72 (1982).

[7]F. Tabakin, Phys. Rev. **137**, B76 (1965); **174**, 1208 (1968).

[8]T. R. Mongan, Phys. Rev. **175**, 1260 (1968); **178**, 1597 (1969).

[9]T. A. Osborn and K. L. Kowalski, Ann. Phys. **68**, 361 (1971).

[10]J. S. Howland, J. Funct. Anal. **22**, 250 (1976).

[11]I. M. Narodetskii, Yad. Fiz. **20**, 1063 (1974) [Sov. J. Nucl. Phys. **20**, 560 (1975)].

[12]B. R. Karlson and E. M. Zeiger, Phys. Rev. D **11**, 939 (1975).

[13]C. Lovelace, in *Strong Interactions and High Energy Physics*, edited by R. G. Moorhouse (Oliver and Boyd, London, 1964).

[14]W. Pogorzelski, *Integral Equations and Their Applications*, edited by I. N. Sneddon and M. Stark (Pergamon, London, 1966), Vol. 1.

[15]E. Schmid and H. Ziegelmann, *The Quantum Mechanical Three-Body Problem*, edited by H. Stumpf (Pergamon, New York, 1974).

The existence of singularities in general relativity despite isolated failures of geodesic focusing

Arvind Borde[a]
Department of Physics, Syracuse University, Syracuse, New York 13244-1130

(Received 4 December 1986; accepted for publication 15 July 1987)

A key ingredient of the singularity theorems of general relativity is that geodesics will focus under certain conditions. While the conditions that are generally imposed are fairly weak and can be expected to hold for most geodesics in any reasonable space-time, the question arises of what the consequences might be if these conditions happen to fail to hold for isolated geodesics. The standard singularity theorems in the literature do not cover such cases. A result is presented here that demonstrates that a singularity will exist even in cases where geodesic focusing does not occur for some geodesics.

I. INTRODUCTION

The aim of the singularity theorems of classical general relativity is to prove that the existence of some appropriate initial condition (e.g., a trapped surface,[1–3] or a reconverging system of geodesics,[3,4] or a compact slice[3–6]) must necessarily lead to a singularity. In order to achieve this, two further classes of assumptions generally have to be made: (I) global conditions of varying reasonableness are often imposed (e.g., global hyperbolicity,[1,2] or the absence of causality violations[3,4]); and (II) some restrictions are placed on the curvature of space-time in order to guarantee that geodesics will have a tendency to converge toward each other and focus. The restrictions on the curvature that are generally imposed are the timelike or null convergence conditions (which follow if Einstein's equation holds and if the geodesics encounter matter with non-negative energy density and pressure)[7] and in some cases the generic condition (which requires that every geodesic feel a nonzero gravitational tidal force).[3,7]

The question then arises of how crucially the existence of a singularity depends on these assumptions. Though the issue is not yet conclusively settled, there is some evidence that assumptions of class I are not crucial[8,9]; this evidence is strengthened by the fact that there are closed universe singularity theorems[4–6] that do not make any such assumptions. On the other hand, class II assumptions enter into the proofs of the singularity theorems in a nontrivial way. Indeed one would not expect singularities to exist, even given an appropriate initial condition, if there are large violations of these conditions. For the singularity theorems are a reflection of the physical fact that gravitation is a universally attractive phenomenon. This physical fact finds its expression in the singularity theorems through the tendency of geodesics to converge and focus. If geodesics do not focus, we are unlikely to obtain a singularity. Thus, for example, de Sitter space-time is nonsingular despite having a compact slice: it does not obey the timelike convergence condition and, consequently, some classes of initially converging timelike geodesics eventually reexpand without coming to a focus. Another example is provided by compactifying spatial sections of Minkowski space-time (so that their topology is $S^1 \times S^1 \times S^1$). In this case the convergence conditions do hold but the generic condition does not, and this space-time too avoids being singular. So, for class II assumptions the appropriate question appears to be not whether they can be eliminated altogether, but, instead, how much they can be weakened while still yielding a singularity.

There are two types of situations that might need weaker class II assumptions than are usually made. The first type of situation is one where there are extended regions in which the convergence conditions do not hold and in which, as a consequence, geodesics will tend to defocus. This happens in some types of cosmological models (e.g., inflationary models) that have been seriously discussed.[10,11] Many of the situations of this type are covered by work that replaces the standard pointwise convergence conditions in the theory of geodesic focusing by weaker integral conditions.[10,11] It follows from this work that limited violations of the energy conditions, even over extended regions, will not affect the existence of singularities. The second type of situation is one where isolated geodesics escape focusing effects. This might happen in situations where most geodesics obey the integral convergence conditions mentioned above, but some isolated geodesics just fail to do so, or if some isolated geodesics do not feel gravitational tidal forces. This is, then, the principal question being considered in this paper: Can isolated failures of geodesic focusing lead to the avoidance of a singularity? The answer that will be obtained is that they cannot. (In the statement and discussion of this result that is presented below, the notation and conventions that will be used are those of Ref. 7.)

The importance of the question being considered here is illustrated by examining the most comprehensive of the singularity theorems, the Hawking–Penrose theorem.[3,7] There, it is shown that space-time cannot be casual geodesically complete if (a) there is an achronal set S such that $E^-(S)$ [or $E^+(S)$] is compact; (b) there are no closed timelike curves; and (c) every complete causal geodesic contains a pair of conjugate points. In this result condition (a) provides the appropriate initial condition. It will hold in a number of different situations. If S is a compact slice then $E^-(S) = S$. And, in a geodesically complete space-time, if S is a past

[a] Present address: Department of Mathematics, Long Island University–Southampton Campus, Southampton, New York 11968.

 0022-2488/87/112683-02$02.50

trapped surface or a point p whose past null cone reconverges and if an averaged convergence condition holds on the null generators of $E^-(S)$, then also $E^-(S)$ will be compact.[11] In connection with reconverging null cones it should be noted that observations of the microwave background provide direct evidence that our own past null cone does in fact reconverge.[3,7] Condition (b) of the theorem is a reasonable causality assumption. (As was said earlier: there is some evidence—even though it is not conclusive—that such assumptions are not crucial to the existence of singularities.) Condition (c) roughly requires that every causal geodesic have a pair of points on it at which it is intersected by "infinitesimally close" nearby geodesics. It will hold in a geodesically complete space-time if the generic condition holds and if the convergence conditions (either the pointwise conditions[7] or the weaker integral conditions[11]) hold everywhere. Since the theorem establishes the existence, under these conditions, of only a single incomplete geodesic it becomes important to know if the failure of condition (c) for some geodesics will affect the existence of the singularity. The fact that it cannot, as long as the failure is confined to isolated geodesics, follows from the result in the next section.

II. A SINGULARITY THEOREM

Theorem: A space-time cannot be past causal geodesically complete if (A) there is an achronal set S such that $E^-(S)$ is compact; (B) strong causality holds (i.e., there are no "almost closed" causal curves); and (C) for every $p \in S$ the following is true: any set U of past-complete causal geodesics whose initial (future) end points form an open neighborhood of p in $E^-(s)$ contains a member with a pair of conjugate points to the past of $E^-(s)$.

Note that the theorem can also be stated in a time-reversed way so as to give a future singularity (as we want for gravitational collapse). It has been stated here in a way that stresses its similarity to the Hawking–Penrose (HP) theorem. Condition (A) here is the same as (a) in the Hawking–Penrose result and (B) is a slight strengthening of (b) there. Condition (C) is analogous to (c) of the HP theorem, but is both weaker than it as well as stronger than it in important respects. It is weaker in that geodesics need not have conjugate points as long as neighboring ones do. This achieves the purpose of this paper. But the requirement that the conjugate points lie to the past of $E^-(S)$ is stronger than requiring that a pair of conjugate points lie somewhere on a geodesic. Such conditions (i.e., that a pair of conjugate points lie on one side of a given point on a geodesic) have been discussed before,[7,11] most notably in a theorem of Hawking[4,7] which assumes in effect that there exists a point p such that, if space-time is past complete, then every past-directed causal geodesic from p has a point on it conjugate to p in $J^-(p)$. In cosmological models of the Robertson–Walker type, such conditions have been explicitly shown to hold (pp. 356–358 of Ref. 7). And they are likely to hold in any situation where the matter density does not drop off too rapidly in the direction of time that is of interest. This will happen in expanding cosmologies (past direction) or gravitational collapse (future). (Despite these arguments about the reasonableness of condition (C), it would be interesting to try to further weaken it. One possible weakening would be a statement of the form "almost every (in some suitable sense) causal geodesic contains a pair of conjugate points." Such an assumption would be a true generalization of the HP generic condition in that it is not tied to the set S at all. Less weak than this statement [but still weaker than condition (C)] would be a condition of the form "if γ is a past-complete geodesic with future end point p then every neighborhood of it contains a causal geodesic with a pair of conjugate points.") In return for this stronger requirement, though, we get two additional pieces of information: the singularity is localized to the past (as it is also in Hawking's theorem) and (as will be seen from the proof of the theorem) as long as (A) and (B) hold there will be an infinite number of past incomplete geodesics. The proof of the result follows the Hawking–Penrose proof closely.

Proof: (A) and (B) imply that $D^-(E^-(S))$ contains a past inextendible timelike curve γ with future end point p on S.[3,7] Let b_i be a sequence of points on γ such that (i) $b_{i+1} \in I^-(b_i)$ and (ii) no compact segment of γ contains an infinite number of the b_i. Let $0 \subset I^+(b_1) \cap E^-(S)$ and let $q \in 0$. Between each b_i and q there will be a timelike geodesic segment of length greater than or equal to that of any causal curve between b_i and q.[7] This sequence of maximal geodesic segments will have a limit geodesic through q denoted by γ_q. Let $U = \{\gamma_q; q \in 0\}$. If the space-time were past complete, then by (C) there would exist a past-directed geodesic μ in U with a pair of conjugate points to the past of 0. The existence of these conjugate points contradicts the fact that μ is a limit geodesic to a sequence of maximal geodesic segments, exactly as in the Hawking–Penrose theorem.

ACKNOWLEDGMENTS

The question discussed here was put to me by Rafael Sorkin and I would like to thank him for many stimulating discussions on geodesic focusing and singularity theorems. I would also like to thank Roger Penrose for a suggestion on how to word the statement of the theorem presented here and Deborah Donahue for typing the manuscript.

Financial support was provided by NSF Grants No. PHY 8318350 and No. PHY 8310041.

[1] R. Penrose, Phys. Rev. Lett. **14**, 57 (1965).
[2] S. W. Hawking, Phys. Rev. Lett. **15**, 689 (1966).
[3] S. W. Hawking and R. Penrose, Proc. R. Soc. London Ser. A **314**, 529 (1970).
[4] S. W. Hawking, Proc. R. Soc. London Ser. A **300**, 187 (1967).
[5] R. P. Geroch, Phys. Rev. Lett. **17**, 445 (1966).
[6] A. Borde, Classical Quantum Gravit. **2**, 589 (1985).
[7] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U. P., Cambridge, MA, 1973).
[8] F. Tipler, Ann. Phys. (NY) **108**, 1 (1977).
[9] A. Borde, Ph.D. dissertation, SUNY at Stony Brook, 1982.
[10] F. Tipler, Phys. Rev. D **17**, 2521 (1978).
[11] A. Borde, Classical Quantum Gravit. **4**, 343 (1987).

Strong cosmic censorship and the strong curvature singularities

Andrzej Królak[a),b)]
Department of Applied Mathematics and Astronomy, University College Cardiff, Cardiff CF1 1XL, United Kingdom

(Received 15 April 1987; accepted for publication 1 July 1987)

Conditions are given under which any asymptotically simple and empty space-time that has a partial Cauchy surface with an asymptotically simple past is globally hyperbolic. It is shown that this result suggests that the Cauchy horizons of the type occurring in Reissner–Nordström and Kerr space-times are unstable. This in turn gives support for the validity of the strong cosmic censorship hypothesis.

I. INTRODUCTION

It has been conjuctured by Penrose[1] that singularities arising in gravitational collapse cannot be visible by observers situated at infinity. This conjecture is known as the weak cosmic censorship hypothesis. Further reflection on this matter has led Penrose to conclude that it makes little difference in terms of predictability whether the singularity is visible from infinity or it can be seen by observers located some finite distance from it.[2] Thus if cosmic censorship is true it should forbid the appearance of locally naked singularities, that is, singularities that lie to the future of some point on an observer's world line and to the past of another point on the same world line. The precise mathematical condition that excludes the locally naked singularities is the demand that space-time be globally hyperbolic.[2] *We shall say that strong cosmic censorship holds if space-time is globally hyperbolic.*

The present author has laid out conditions under which the weak cosmic censorship holds,[3,4] the main consideration of which was the condition that all singularities arising in gravitational collapse should be of strong curvature type (for a definition see the next section). In the next section we shall give a set of conditions under which any weakly asymptotically simple and empty space-time with a partial Cauchy surface with an asymptotically simple past is globally hyperbolic. We shall show that this result indicates that Cauchy horizons of the type occurring in Reissner–Nordström and Kerr space-times are unstable.

II. CONDITIONS FOR GLOBAL HYPERBOLICITY OF AN ASYMPTOTICALLY FLAT SPACE-TIME

Before we shall state and prove our main theorem we shall explain the basic notions used in this theorem and we shall give some definitions.

A commonly accepted precise definition of an asymptotically flat space-time is contained in the notion of a weakly asymptotically simple and empty space-time.[5] The existence of a partial Cauchy surface $\mathscr{S}$ with an asymptotically simple past in such a space-time ensures that singularities occurring in space-time arise from an initially regular state and that any Cauchy horizons arising to the future of $\mathscr{S}$ are not artifacts of a bad choice of the initial surface but rather are a result of the genuine breakdown of global hyperbolicity. For example, a partial Cauchy surface in an asymptotically simple and empty space-time (a globally hyperbolic space) with an edge on past null infinity $\mathscr{I}^-$ or future null infinity $\mathscr{I}^+$ would contain a Cauchy horizon.

Definition 1: A future-endless nonspacelike geodesic λ is said to terminate in a strong curvature singularity in the future, if for every point p on λ the expansion Θ of the future-directed nonrotating congruence of geodesics from p containing λ becomes negative somewhere on λ.

A suggestion that all singularities arising in physically realistic space-times are of strong curvature type has been put forward by this author[6] and independently by Tipler *et al.*[7] In our opinion the above definition is a precise geometrical description of the singularity arising as a result of reaching by the gravitational field the point of no return beyond which only its further, unbounded increase in strength is possible. All other singularities are in a certain sense artificial originating from some special conditions imposed on space-time, e.g., symmetry, Petrov type, special initial conditions.

The above definition is a modification of Tipler's original one.[8]

Definition 2: Tipler's condition is said to hold on a future Cauchy horizon H^+ if on every past-endless null geodesic generator $\lambda(u)$ of H^+ either

(i) $\lim_{u\to c}(\inf R_{ab}k^a k^b) > 0$

or

(ii) $\lim_{u\to c}(\inf|k^c k^d k_{[a}C_{b]cd[e}k_{f]}|) > 0,$

where c is the past limit of the affine parameter u along λ and k^a is the tangent vector to λ.

The conditions in the above definition were introduced by Tipler.[9] They express the idea that Cauchy horizons forming in a space-time result from a certain arrangement of a sufficient amount of matter or gravitational radiation and not, for example, from some identifications made in an otherwise regular space-time.

[a)] SERC Postdoctoral research assistant.
[b)] On leave of absence from the Mathematical Institute, Polish Academy of Sciences, Warsaw, Poland.

Definition 3: The simplicity condition is said to hold on a future Cauchy horizon $H^+(\mathscr{S})$, $\mathscr{S}$ is a partial Cauchy surface, if there exists a sequence (λ_i) of null geodesics past-endless in the domain of dependence $D(\mathscr{S})$ of $\mathscr{S}$ each generating an achronal set and such that the sequence (λ_i) has a past-endless null geodesic generator λ of $H^+(\mathscr{S})$ as its limit curve.

The simplicity condition need not hold for an arbitrary Cauchy horizon. This is because a past-endless null geodesic may encounter caustics to the past of which it can no longer generate an achronal set. Nevertheless we shall show below that the simplicity condition holds for the type of Cauchy horizons occurring in the Reissner–Nordström or Kerr space-times.

We can now state our theorem.

Theorem 1: A weakly asymptotically simple and empty space-time $(\mathscr{M}, g)$ containing a partial Cauchy surface $\mathscr{S}$ with an asymptotically simple past is globally hyperbolic if the following conditions are satisfied.

(1) $R_{ab}k^a k^b > 0$ for every null vector k^a.

(2) The strong causality condition holds.

(3) If the future Cauchy horizon $H^+(\mathscr{S})$ is not empty then both Tipler's condition and the simplicity condition hold.

(4) Each past-incomplete null geodesic terminates in a strong curvature singularity in the past.

Proof: Suppose that space-time $(\mathscr{M}, g)$ is not globally hyperbolic. Then either $H^+(\mathscr{S})$ or $H^-(\mathscr{S})$ is not empty. Since space-time is assumed to have an asymptotically simple past we have $D^-(\mathscr{S}) = J^-(\mathscr{S})$ by definition and therefore $H^+(\mathscr{S}) \neq \varnothing$.

Let λ be any past-endless null geodesic generator of $H^+(\mathscr{S})$. By Tipler's conditions the generator λ cannot be past complete. If it were complete by Lemma 9 and the proof of Proposition 5 in Ref. 9, λ would contain an infinite number of conjugate points. Thus by Ref. 5, Proposition 4.5.12, points of $H^+(\mathscr{S})$ could be joined by timelike curves. This is impossible as the set $H^+(\mathscr{S})$ is achronal.

Let (λ_i) be the sequence of past-endless null geodesics in int $D(\mathscr{S})$ given in Definition 3. Let p_i be a sequence of points, each $p_i \in \lambda_i$ such that (p_i) has a limit point on the generator λ of $H^+(\mathscr{S})$. Each (λ_i) when followed into the past must eventually intersect $\mathscr{S}$ and consequently it must have a past end point on $\mathscr{J}^-$ because $\mathscr{S}$ has an asymptotically simple past. Thus each (λ_i) is past complete. Consequently the expansion Θ_i on (λ_i) of the congruence of past-directed null geodesics from p_i cannot become negative otherwise, by the Raychaudhuri equation and condition (1), there would be a point conjugate to p_i on λ_i. This is impossible since each λ_i generates an achronal set. By continuity, the expansion Θ on λ of the past-directed congruence of null geodesics from p is also non-negative. However, as λ is a past-endless generator of $H^+(\mathscr{S})$, by the first part of this proof it is past-incomplete. Therefore by condition (4) it terminates in a strong curvature singularity in the past. Consequently, by Definition 1, Θ_i must become negative. This is a contradiction. Thus space-time must be globally hyperbolic. □

In Figs. 1 and 2 we have shown two types of Cauchy horizons that can arise in an asymptotically flat space-time as a result of the occurrence of a space-time singularity to the future of a partial Cauchy surface $\mathscr{S}$. The Cauchy horizon depicted in Fig. 1 is of the type occurring in the Rissner–Nordström space-time for $m^2 > e^2$ or Kerr space-time for $m^2 > a^2$, where m is the mass of the black hole, a its angular momentum, and e its charge. We can distinguish the two types of Cauchy horizons by the following criterion. For the Cauchy horizon in Fig. 1 there exists a point p on $H^+(\mathscr{S})$ such that in the chronological past $I^-(p, \bar{\mathscr{M}})$ of p in $\bar{\mathscr{M}}$ (where $\bar{\mathscr{M}} = \mathscr{M} \cup \mathscr{J}^+ \cup \mathscr{J}^-$), there exists a future-endless null geodesic generator γ of $\mathscr{J}^-$ entirely contained in $I^-(p, \bar{\mathscr{M}})$ whereas for the Cauchy horizon in Fig. 2, for any point p on the horizon all the generators of $\mathscr{J}^-$ leave $I^-(p, \bar{\mathscr{M}})$ when maximally extended to the future.

In the following lemma we shall show that if the above criterion is satisfied then the simplicity condition holds.

Lemma 1: Let $\mathscr{S}$ be a partial Cauchy surface with an asymptotically simple past. Suppose that the future Cauchy horizon $H^+(\mathscr{S})$ is not empty and suppose that there exists a point p on $H^+(\mathscr{S})$ such that p is not a future end point of a generator of $H^+(\mathscr{S})$ and such that $I^-(p, \bar{\mathscr{M}})$ contains a future-endless null geodesic generator γ of $\mathscr{J}^-$ then simplicity condition holds on $H^+(\mathscr{S})$.

Proof: Consider a sequence of points (p_i) in int $D(\mathscr{S})$ converging to the point p such that for each i, $p_{i+1} \in I^+(p_i)$. Consider the boundary $\dot{J}^-(p_i, \bar{\mathscr{M}})$, where $\bar{\mathscr{M}} = \mathscr{M} \cup \mathscr{J}^+ \cup \mathscr{J}^-$. Since the surface $\mathscr{S}$ has an asymptotically simple past, by the time reverse of the argument contained in Lemma 6.9.3 of Ref. 5 each generator of $\mathscr{J}^-$ and therefore the generator γ intersects $\dot{J}^-(p_i, \bar{\mathscr{M}})$ exactly once. Let us denote the point of intersection by q_i and the generator of $\dot{J}^-(p_i, \bar{\mathscr{M}})$ joining q_i and p_i by λ_i. Thus we have a sequence of null geodesics (λ_i) such that each member of the sequence generates an achronal set $(\dot{J}^-(p_i, \bar{\mathscr{M}}))$ and such that p is a limit point of (λ_i). Each λ_i is past endless in $\mathscr{M}$.

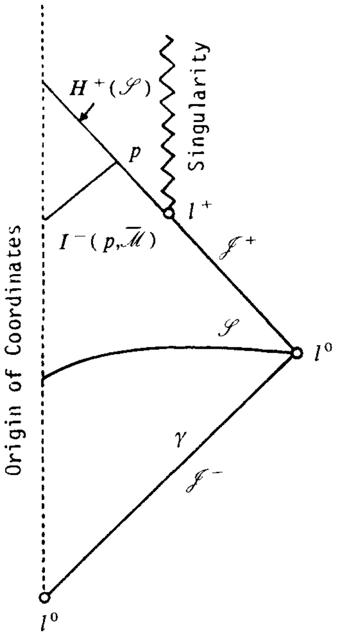

FIG. 1. Cauchy horizon of the type occurring in the Reissner–Nordström space-time for $m^2 > e^2$ or Kerr space-time for $m^2 > a^2$. A generator γ of $\mathscr{J}^-$ is contained in the past of a point p on the horizon.

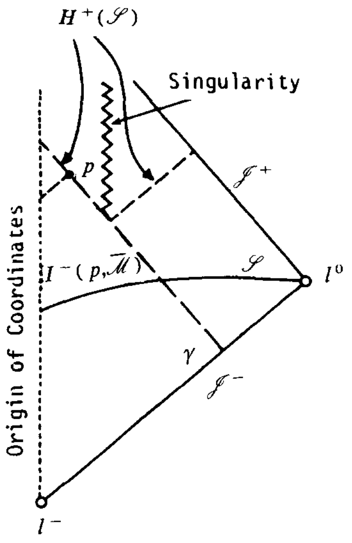

FIG. 2. An example of a Cauchy horizon such that for any point p on the horizon all the generators of $\mathscr{I}^-$ leave the past at p when extended into the future.

Hence by time reverse of Lemma 6.2.1 there is a nonspacelike curve λ which is past endless in $\mathscr{M}$ and which is a limit curve of (λ_i). Since members of the sequence (λ_i) are null geodesics, λ is a null geodesic as well. We shall show that λ is a generator of $H^+(\mathscr{S})$. This will complete the proof of the lemma.

We first show that $q_{i+1} \in J^+(q_i)$. If not then $q_{i+1} \in J^-(q_i) - \{q_i\}$. Since $q_i \in \dot{J}^-(p_i, \mathscr{M})$ and $p_{i+1} \in I^+(p_i)$ then $p_{i+1} \in I^+(q_{i+1}, \mathscr{M})$. This is a contradiction since by construction $q_{i+1} \in \dot{J}^-(p_{i+1}, \mathscr{M})$ and therefore points q_{i+1} and p_{i+1} lie in the same achronal set $\dot{J}^-(p_{i+1}, \mathscr{M})$, and thus cannot be joined by a timelike curve. Next we show that the sequence (q_i) has no limit point. If (q_i) had a limit point q then $q \in \gamma$ since by construction for each i, $q_i \in \gamma$. Since $\gamma \subset I^-(p, \mathscr{M})$ then $q \in I^-(p, \mathscr{M})$. Thus we could find a point p_I of the sequence (p_i) such that $q \in I^-(p_I, \mathscr{M})$. Since for each i, $q_{i+1} \in J^+(q_i)$, all the members of the sequence (q_i) are in $J^-(q)$. Hence $q_I \in J^-(q)$ and therefore $q_I \in I^-(p_I, \mathscr{M})$. This is a contradiction as by construction $q_I \in \dot{J}^-(p_I, \mathscr{M})$. Finally suppose that the limit curve λ of the sequence (λ_i) is not a past-endless null geodesic generator of $H^+(\mathscr{S})$ but some other past-endless null geodesic through p. Since λ is not a generator of $H^+(\mathscr{S})$ it must enter int $D(\mathscr{S})$ to the past of p, intersect $\mathscr{S}$, and finally reach $\mathscr{I}^-$ when maximally extended into the past since the partial Cauchy surface $\mathscr{S}$ has an asymptotically simple past. The point of intersection of λ with $\mathscr{I}^-$ would be the limit point of the sequence (q_i). This is a contradiction as we have shown that (q_i) has no limit point. □

Thus by the above lemma the second part of condition (3) of Theorem 1 is fulfilled for the case of Cauchy horizon occurring in Reissner–Nordström space-time for $m^2 > e^2$ and Kerr space-time for $m^2 > a^2$. Tipler's condition is not satisfied in the Reissner–Nordström or Kerr space-times. However, it was shown by Tipler[10] that if there is an outgoing spherically symmetric radiation of arbitrary small density in some neighborhood of the event horizon in Reissner–Nordström space-time then Tipler's condition holds. This gives us reason to believe that in perturbed Reissner–Nordström or Kerr space-time Tipler's condition will be fulfilled. Consequently, Theorem 1 above suggests that Cauchy horizons in Reissner–Nordström or Kerr space-times will not appear in generic space-times. This gives evidence for the validity of the strong cosmic censorship hypothesis.

ACKNOWLEDGMENTS

I am grateful to the Department of Applied Mathematics and Astronomy, University College Cardiff, and in particular to Professor B. F. Schutz for their hospitality. I would like to thank Dr. St. Bazanski, Professor C. J. S. Clarke, and W. Rudnicki for helpful discussions and D. Baxter for drawing the figures.

[1]R. Penrose, Riv. Nuovo Cimento **1**, 252 (1969).
[2]R. Penrose in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U. P., Cambridge, 1979), p. 581.
[3]A. Królak, Gen. Relativ. Gravit. **15**, 99 (1983).
[4]A. Królak, J. Class. Quantum Gravit. **3**, 267 (1986).
[5]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1973).
[6]A. Królak, Master's thesis, Warsaw University, 1978 (unpublished).
[7]F. J. Tipler, C. J. S. Clarke, and F. G. R. Ellis, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York 1980).
[8]F. J. Tipler, Phys. Lett. A **64**, 8 (1977).
[9]F. J. Tipler, Ann. Phys. (NY) **108**, 1 (1977).
[10]F. J. Tipler, Phys. Rev. D **16**, 3359 (1977).

Commutation properties of cyclic and null Killing symmetries

L. B. Szabados
Central Research Institute for Physics, H-1525 Budapest 114, P. O. Box 49, Hungary

(Received 12 February 1987; accepted for publication 1 July 1987)

In a space-time admitting cyclic and nonspacelike Killing symmetries the commutation properties of the Killing vectors are examined. It is shown that cyclic and null Killing vectors can be noncommuting only if a Killing vector of stationarity is also admitted. Two consequences of this commutativity are also discussed.

I. INTRODUCTION

Stationary, axisymmetric space-times have distinguished importance in general relativity, e.g., the final state of black holes is thought of as stationary and axisymmetric.[1] It is usually assumed that the Killing vectors of stationarity and axisymmetry commute. In fact, Carter[2] has shown that, without loss of any generality, this can always be assumed.

Axisymmetric space-times with null Killing vectors may also have physical significance; e.g., certain *pp* waves[3] or the Lukács–Perjés–Sebestyén solution[4] (which describes the gravitational field of a zero-mass, spinning charged particle) have these symmetries. Recently Lessner[5] has proposed certain axisymmetric vacuum solutions, admitting null Killing symmetry, of the five-dimensional Einstein equations as models of extended massless particles. The commutation of these Killing vectors is also assumed. Unfortunately, this commutation property does not follow from Carter's theorem.

In the present paper we generalize Carter's theorem. No fixed point is needed, so the axial symmetry is weakened to cyclic symmetry; and the timelike Killing symmetry is replaced by a nonspacelike one. We show that in a cyclically symmetric space-time, admitting a null Killing vector field, the two Killing vectors must commute, unless otherwise, in addition, the space-time has to admit a timelike Killing symmetry, too. Finally, based on this commutation property, we give a sufficient condition on a cyclically and null Killing symmetric space-time to be in Kundt's class[3] and it is shown that in space-times describing axial symmetric *pp* waves the null Killing vector must be orthogonal to the orbits of axial symmetry.

By space-time we mean a smooth, paracompact four-dimensional manifold M endowed with a Lorentzian metric,[1] but we do not use any field equation.

II. CYCLICALLY SYMMETRIC SPACE-TIME WITH NONSPACELIKE KILLING SYMMETRY

Space-time (M,g) is said to be cyclically symmetric[2] if there is a smooth map σ: $SO(2)\times M\to M$: $(\varphi,p)\mapsto\sigma(\varphi,p)$ for which each of the following conditions holds: (1) $\forall\varphi\in SO(2)$ the map $\sigma(\varphi):M\to M:p\mapsto\sigma(\varphi,p)$ is an isometry of (M,g); (2) $\forall\varphi',\varphi\in SO(2)$, $\sigma(\varphi)\circ\sigma(\varphi')=\sigma(\varphi+\varphi')$; (3) if $\sigma(\varphi)=\mathrm{Id}_M$ then $\varphi=0$ [i.e., SO(2) acts on M effectively]; and (4) the vector $X_p:=(\partial/\partial\varphi)_{\sigma(\varphi,p)|\varphi=0}$ is spacelike $\forall p\in M$.

One can define the orbit through p as $O(p):=\{\sigma(\varphi,p)\,|\,\varphi\in SO(2)\}$, and p is said to be a fixed point if $O(p)=\{p\}$. It is easy to show that p is a fixed point iff $X_p=0$, and if p is not a fixed point then there is a diffeomorphism of SO(2) onto $O(p)$ and so X, defined pointwise by $p\mapsto X_p$, is a smooth vector field. Here X will be called a cyclic Killing vector field.

Proposition 1: Let (M,g) be cyclically symmetric with SO(2) action σ and cyclic Killing vector field X, and let K be a nowhere vanishing future directed smooth nonspacelike Killing vector field on M. Then the vector field $\widetilde{K}$, defined pointwise by

$$\widetilde{K}_p:=\frac{1}{2\pi}\int_0^{2\pi}(\sigma(\varphi)_*K)_p\,d\varphi,\quad p\in M,$$

is a future directed nowhere vanishing smooth nonspacelike Killing vector field, which is invariant under the action σ; i.e., $[X,\widetilde{K}]=0$.

Proof: Since $\forall\varphi\in SO(2)$, $\sigma(\varphi)$ is an isometry, thus $\sigma(\varphi)_*K$ is a nowhere zero smooth nonspacelike Killing vector field. Here M is time oriented, therefore $\sigma(\varphi)_*K$ is also future directed. Consequently, $\widetilde{K}$ is a nowhere vanishing future directed nonspacelike smooth Killing vector field. $\forall\varphi\in SO(2)$

$$\sigma(\varphi)_*\widetilde{K}=\frac{1}{2\pi}\int_0^{2\pi}(\sigma(\varphi+\varphi')_*K)d\varphi'$$

$$=\frac{1}{2\pi}\int_0^{2\pi}(\sigma(\varphi')_*K)d\varphi'=\widetilde{K},$$

i.e., $\widetilde{K}$ is invariant under the action σ. However, because of Corollary 1.8 and 1.11 of Ref. 6, this is equivalent to $[X,\widetilde{K}]=0$. □

Recall that the space-time is said to be stationary if it admits a nowhere vanishing smooth timelike Killing vector field.

Corollary: Let (M,g) be stationary and cyclically symmetric with cyclic Killing vector field X. Then there is a future directed smooth timelike vector field V which commutes with X.

This statement is a generalization of Carter's theorem[2]: it guarantees the existence of a timelike Killing vector field commuting with that of the cyclic symmetry in every stationary cyclically symmetric space-time, even in the presence of wire singularity. The existence of fixed points of σ is not needed; moreover, no restriction is required for the dimension of the space-time: it can be used for higher-dimensional Lorentzian geometries (e.g., in Kaluza–Klein theories) too.

 0022-2488/87/112688-04$02.50

If K is timelike then $\tilde{K}$ must be timelike too. If, however, K is nonspacelike or null, then $\tilde{K}$ may be timelike on an (open) set and null on its complement. In the rest of this section, where the space-time is assumed to be cyclically and nonspacelike Killing symmetric with SO(2) action σ and Killing fields X and K, respectively, the causal character of $\tilde{K}$ will be considered in fixed points of σ and along orbits diffeomorphic with SO(2) as well.

Proposition 2: If p is a fixed point of σ, then $\tilde{K}$ is null at p iff K is null at p and $[X,K]$ vanishes at p.

Proof: Here p is fixed, thus $\forall\varphi\in\mathrm{SO}(2)$, $\sigma(\varphi,p)=p$ and $\sigma(\varphi)_*:T_pM\to T_pM$. Here $\tilde{K}$ can be null at p only if K is null at p and there is a positive smooth function $f(\varphi)$ such that $\sigma(\varphi)_*K_p=f(\varphi)K_p$. This implies $f(\varphi+\varphi')=f(\varphi)f(\varphi')$, $\forall\varphi,\varphi'\in\mathrm{SO}(2)$. Its solution is $f(\varphi)=\exp(f'(0)\varphi)$. But $f(\varphi)=f(2\pi+\varphi)$ must hold, thus $f'(0)=0$ and $\sigma(\varphi)_*K_p=K_p$; i.e., $[X,K]$ vanishes at p.

Conversely, if K_p is null at p and $[X,K]$ vanishes at p then $\sigma(\varphi)_*K_p=K_p$ for $\forall\varphi\in\mathrm{SO}(2)$ and $\tilde{K}_p=K_p$ is null. □

Now suppose p is not a fixed point of σ. Then $\forall\, q\in O(p)$ there is a unique element $\psi\in\mathrm{SO}(2)$ for which $q=\sigma(\psi,p)$. The following statement gives necessary and sufficient conditions that guarantee $\tilde{K}$ being null along the orbit $O(p)$.

Proposition 3: The vector field $\tilde{K}$ is null along the orbit $O(p)$ iff K is null along $O(p)$ and there is a smooth positive function $\Phi(\psi)$ for which $[X,K]_q=-\Phi(\psi)K_q$, $q=\sigma(\psi,p)$. [For such a $\Phi(\psi)$ the integral $\int_0^{2\pi}\Phi(\psi)d\psi$ is necessarily zero.]

Proof: The vector field $\tilde{K}$ can be null at $q\in O(p)$ only if K is null all along $O(p)$ and there is a smooth positive function $f(\psi,\varphi)$ for which

$$(\sigma(\varphi)_*K)_q=f(\psi,\varphi)K_q\,.$$

This implies

$$f(\psi,\varphi+\varphi')=f(\psi,\varphi')f(\psi-\varphi',\varphi)\,, \tag{1}$$

$\forall\varphi,\varphi'\in\mathrm{SO}(2)$ and $q\in O(p)$. Let $F(\psi,\varphi):=\ln f(\psi,\varphi)$ and, denoting the derivative of F with respect to its first and second argument by F_1 and F_2, respectively, one obtains

$$F_2(\psi,\varphi+\varphi')=F_2(\psi,\varphi')-F_1(\psi-\varphi',\varphi)\,, \tag{2}$$

$$F_2(\psi,\varphi+\varphi')=F_2(\psi-\varphi',\varphi)\,. \tag{3}$$

The solution of Eq. (3) must have the form

$$F_2(\psi,\varphi)=\Phi(\psi-\varphi)\,, \tag{4}$$

where Φ is a smooth function. Using this expression, Eq. (2) yields

$$F_1(\psi,\varphi)=\Phi(\psi)-\Phi(\psi-\varphi)\,. \tag{5}$$

The integrability conditions for the system of partial differential equations (4) and (5) hold identically, and its solution is

$$F(\psi,\varphi)=\int_{\psi-\varphi}^{\psi}\Phi(u)du+F_0\,. \tag{6}$$

Substituting (6) into Eq. (1) one obtains $F_0=0$. Since f is periodic, i.e., $f(\psi,\varphi+2\pi)=f(\psi,\varphi)$, it follows that

$$\int_{\psi}^{\psi+2\pi}\Phi(u)du=0\,,$$

$\forall\psi\in[0,2\pi]$, from which $\Phi(\psi+2\pi)=\Phi(\psi)$. This condition guarantees $f(\psi,\varphi)=f(\psi+2\pi,\varphi)$, too. According to Proposition 1, the vector field $\tilde{K}$ is invariant under the action σ, thus

$$\begin{aligned}0&=2\pi[X,\tilde{K}]\\&=\int_0^{2\pi}f(\psi,\varphi)d\varphi\,[X,K]+\frac{d}{d\psi}\int_0^{2\pi}f(\psi,\varphi)d\varphi\,K\\&=\int_0^{2\pi}f(\psi,\varphi)d\varphi\,[X,K]\\&\quad+\int_0^{2\pi}f(\psi,\varphi)(\Phi(\psi)-\Phi(\psi-\varphi))d\varphi\,K\,.\end{aligned}$$

But

$$\begin{aligned}&\int_0^{2\pi}f(\psi,\varphi)\Phi(\psi-\varphi)d\varphi\\&\quad=-\int_0^{2\pi}\frac{\partial f}{\partial\varphi}d\varphi=-f(\psi,2\pi)+f(\psi,0)=0\,,\end{aligned}$$

thus

$$[X,K]=-\Phi(\psi)K\,.$$

Conversely, if there is a smooth function $\Phi(\psi)$ for which $[X,K]=-\Phi(\psi)K$, then in the coordinate system (x^0,x^1,x^2,ψ) adapted to X, where the orbits are given by $x^0,x^1,x^2=\mathrm{const}$,

$$K^a_q=K^a_p\exp\Big(-\int_0^{\psi}\Phi(u)du\Big)\,.$$

But $p=\sigma(2\pi,p)$, thus $K^a_{\sigma(2\pi,p)}$ must be equal to K^a_p. This implies $\int_0^{2\pi}\Phi(u)du=0$. The action of $\sigma(\varphi)_*$ on K can be calculated in the coordinate system (x^0,x^1,x^2,ψ):

$$\begin{aligned}(\sigma(\varphi)_*K)^a_q&=K^a_{\sigma(-\varphi,q)}\\&=K^a_p\exp\Big(-\int_0^{\psi-\varphi}\Phi(u)du\Big)\\&=K^a_q\exp\int_{\psi-\varphi}^{\psi}\Phi(u)du\,.\end{aligned}$$

Thus if K is null along $O(p)$ then

$$\tilde{K}_q=\frac{1}{2\pi}\int_0^{2\pi}\Big(\exp\int_{\psi-\varphi}^{\psi}\Phi(u)du\Big)d\varphi\,K_q$$

is also null for $\forall q\in O(p)$. □

Corollary: Let $q=\sigma(\psi,p)$ be a point of $O(p)$, where $g(X,K)\neq0$. Then $\tilde{K}$ is null along $O(p)$ iff K is null and $[X,K]=0$ along $O(p)$.

Proof: Since K is a Killing vector field and $\forall\varphi\in\mathrm{SO}(2)$, $\sigma(\varphi)$ is an isometry, $\sigma(\varphi)_*K$ is a null Killing vector field. Thus along $O(p)$ one has

$$\begin{aligned}0&=X^a(\sigma(\varphi)_*K)_{a;b}X^b=f(\psi,\varphi)X^aK_{a;b}X^b+X^aK_a\frac{\partial f}{\partial\psi}\\&=X^aK_af(\psi,\varphi)(\Phi(\psi)-\Phi(\psi-\varphi))\,.\end{aligned}$$

Since X^aK_a is not zero at $q=\sigma(\psi,p)$, $\Phi(\psi)=\Phi(\psi-\varphi)$, $\forall\varphi\in\mathrm{SO}(2)$; i.e., $\Phi=\mathrm{const}$. But the only constant function having zero integral on $[0,2\pi]$ is the zero, thus $f(\psi,\varphi)=1$, $\forall\varphi,\psi\in\mathrm{SO}(2)$; i.e., $[X,K]=0$ along $O(p)$. □

Thus K is null along $O(p)$ iff $[X,K]=0$ and K is null, except the very special case in which X and K are orthogonal all along $O(p)$.

Although, if $X^a K_a = 0$ along $O(p)$, $\Phi(\psi)$ may be nonzero even if both K and $\widetilde{K}$ are null, but then, as the next proposition shows, the whole orbit $O(p)$ lies in the closure of an open set on which $\widetilde{K}$ is timelike.

Proposition 4: If the commutator $[X,K]$ does not vanish at some point $q\in O(p)$, then every neighborhood of each point of $O(p)$ contains a point where $\widetilde{K}$ is timelike.

Proof: If K is timelike at some point of $O(p)$, or if K is null along $O(p)$ but there is no function Φ required in Proposition 3, then $\widetilde{K}$ is timelike on $O(p)$. Thus one can assume that K is null and $[X,K] = -\Phi K$ along $O(p)$ for some smooth function $\Phi(\psi)$.

Let $r\in O(p)$ and W be a neighborhood of r. If K is not null at some point $s\in W$, then $\widetilde{K}$ is timelike on $O(s)$. Then one can assume that K is null on W. If there is a point $s\in W$ where the vector $[X,K]_s$ is not proportional to K_s, then a function Φ_s, required in Proposition 3, could not exist along the orbit $O(s)$, thus $\widetilde{K}$ is timelike on $O(s)$. One can assume therefore that $[X,K]$ is proportional to K on W. It will be shown, however, that K being null on W and $[X,K]$ being proportional to K on W together contradict our hypothesis $[X,K]_q \neq 0$.

If $[X,K]$ is proportional to K, then, because of their smoothness, a function $\widetilde{\Phi}$ exists on W for which $[X,K] = -\widetilde{\Phi}K$ and $\widetilde{\Phi}$ coincides with Φ on the orbit $O(p)$. Here X and K are Killing fields and K is null on W, thus $\widetilde{\Phi}K$ must be a null Killing vector field, therefore

$$0 = (\widetilde{\Phi}K_a)_{;b} + (\widetilde{\Phi}K_b)_{;a} = \widetilde{\Phi}_{;a}K_b + \widetilde{\Phi}_{;b}K_a .$$

Let $s\in W$ and $\{K,L,E_m\}$ be a pseudo-orthonormalized vector base at T_sM. Contracting the above equation with K^aL^b, L^aL^b, and $E^{e}_{m}L^b$ one obtains $\widetilde{\Phi}_{;a}K^a = 0$, $\widetilde{\Phi}_{;a}L^a = 0$, and $\widetilde{\Phi}_{;a}E^{a}_{m} = 0$, respectively; i.e., $d\widetilde{\Phi} = 0$ at s. But s can be chosen arbitrarily, therefore $\widetilde{\Phi} = \Phi_0 = \text{const}$ on W.

The orbit $O(p)$ is compact, so it can be covered by finitely many neighborhoods $W_1,\ldots,W_t$. But, due to the overlappings of the W's, Φ has to be the same constant value Φ_0 all along $O(p)$. This, however, implies $\Phi_0 = 0$, which contradicts the hypothesis $[X,K]_q \neq 0$. □

At the end of this section we review the properties of the two-dimensional orbits. If the action σ has a fixed point p, then a two-dimensional timelike submanifold, called the symmetry axis, can be foliated through p (Ref. 2), and the integral curve of $\widetilde{K}$ through p lies in this axis.

Outside the axis X and $\widetilde{K}$ together constitute a smooth two-dimensional involutive distribution,[7] thus there is a two-dimensional integral submanifold $N(p)$ of $X,\widetilde{K}$ through each nonfixed p. Here $N(p)$ is generated by the orbits $L(q)$ of the one-parameter group action generated by $\widetilde{K}$, and $q\in O(p)$. Since X and $\widetilde{K}$ are commuting Killing fields, all the inner products $g(X,X)$, $g(X,\widetilde{K})$, $g(\widetilde{K},\widetilde{K})$ are constant on $N(p)$. Therefore $N(p)$ is a cylinder with constant circumference and its causal character does not change along the orbits $L(q)$.

If $\widetilde{K}$ is null along $O(p)$, then K is null on $N(p)$ and the integral curves of K and $\widetilde{K}$, lying in $N(p)$, coincide. In this case, because of the Corollary of Proposition 3, and the equality $X(g(X,K)) = g(X,[X,K])$, $g(X,K)$ is constant on $N(p)$, too.

III. CYCLICALLY SYMMETRIC SPACE-TIMES WITH NULL KILLING SYMMETRY

If K is nonspacelike, then, in general, the causal character of $\widetilde{K}$ may vary on M, due to the changing of the causal character of K or of the commutation property of X and K.

Throughout this section K is assumed to be null. Therefore the first possibility above is ruled out, but, as our main theorem states, not the second one.

Theorem: Let (M,g) be cyclically symmetric with cyclic Killing vector field X and SO(2) action σ, and let K be a nowhere vanishing null Killing vector field on M. Then either $[X,K] = 0$ on M, or

$$U := \{\sigma(\varphi,p)\,|\,\varphi\in SO(2),\ p\in M\text{: } \nexists\alpha\in R \text{ for which } [X,K]_p = \alpha K_p\}$$

is a nonempty open set and the Killing vector field $\widetilde{K}$ is timelike on U.

Proof: If $[X,K]$ is not zero at some point $p\in M$, then, as a corollary to Proposition 4, there is an open set V such that $p\in\overline{V}$ and $\widetilde{K}$ is timelike on V. But, as Proposition 3 states, $\widetilde{K}$ is timelike on U and can be timelike only on U. Here U is open and, because of $V\subseteq U$, nonempty. □

This theorem is the main result of the present paper. It states that the Killing vectors X and K can be noncommuting only if the space-time admits an additional timelike Killing symmetry on an open subset of M. Thus if we want to consider space-times only with cyclic and null Killing symmetries, we have to assume they commute, as otherwise stationarity on an open set is also assumed implicitly. (See Note added in proof.)

Recall that a null Killing vector field is always geodesic and its expansion and shear vanish. Thus space-times admitting a null Killing vector field K are classified as the twist ω of K vanishes or not, and in the first case as K is covariantly constant or not.[3] If, however, in addition a cyclic symmetry is also admitted, then a further subclass can be introduced.

Corollary: Let (M,g) be a cyclically and null Killing symmetric space-time with (commuting) Killing vector fields X and K, respectively, and SO(2) action σ.

(1) If $g(X,K) = 0$ throughout M then K is twist-free.

(2) If K is covariantly constant then $g(X,K)$ is constant on M, and, in addition, if σ has a fixed point then $g(X,K) = 0$.

Proof: (1) If p is not a fixed point of σ, then in a neighborhood W of p one can define a unit spacelike smooth vector field Y, being orthogonal to both K and X. This Y is unique up to a sign. The function $x := \{g(X,X)\}^{1/2}$ is nonzero and smooth on W, thus $E_2 := Y$, $E_3 := (1/x)X$ constitute a smooth two-dimensional orthonormal spacelike base field on W, being orthogonal to K. The twist of K can be calculated in this base,

$$\begin{aligned} 2\omega x &= xg(\nabla_{E_2}K,E_3) - xg(\nabla_{E_3}K,E_2) \\ &= -g(K,\nabla_Y X) + g(K,\nabla_X Y) = g(K,[X,Y]) ; \end{aligned}$$

i.e., if Y is Lie propagated along X then K is twist-free.

Let $q\in W$ and let Y' denote the vector field along the orbit $O(p)$, obtained by Lie propagation of the vector Y_q. [Here X is a Killing field, thus, in spite of the fact that $O(p)$ is closed,

Y' is well defined all along $O(p)$.] Then

$$X(g(X,Y')) = Y'^a X_{a;b} X^b + X^a Y'_{a;b} X^b$$
$$= X^a (Y'_{a;b} X^b - X_{a;b} Y'^b) = 0 ,$$
$$X(g(K,Y')) = Y'^a K_{a;b} X^b + K^a Y'_{a;b} X^b$$
$$= Y'^a X_{a;b} K^b + K^a Y'_{a;b} X^b$$
$$= K^a (Y'_{a;b} X^b - X_{a;b} Y'^b) = 0 ,$$
$$\tfrac{1}{2} X(g(Y',Y')) = Y'^a Y'_{a;b} X^b = Y'^a X_{a;b} Y'^b = 0 ;$$

i.e., Y' is the unit spacelike vector field being orthogonal to X and K all along $O(p)$. Thus it coincides with Y; i.e., Y is Lie propagated along X.

(2) Let q be a nonfixed point and let I_q denote the integral of $g(X,K)$ on $O(p)$. As we stated at the end of the previous section, $g(X,K)$ is constant along the orbits $N(q)$, thus $I_q = 2\pi g_q(X,K)$. On the other hand, I_q can be considered as the integral of the closed one-form field K_a on the one-cycle $O(q)$ (K_a is closed; i.e., $K_{[a;b]} = 0$, because K_a is constant):

$$I_q = \int_{O(q)} K .$$

Let q' be an arbitrary point of M. Here M is connected, therefore there is a smooth curve μ: $[0,1] \to M$ from $q = \mu(0)$ to $q' = \mu(1)$. The mapping

$$F\colon [0,1] \times [0,2\pi] \to M\colon (t,\psi) \mapsto \sigma(\psi,\mu(t))$$

is a smooth homotopy between the orbits $O(q)$ and $O(q')$, thus $I_q = I_{q'}$, which implies $g(X,K) = \text{const}$. If there is a fixed point p, then q' can be chosen to be p, therefore, because of $O(p) = \{p\}$, every orbit $O(q)$ is homotopic to zero and consequently $g(X,K) = 0$. □

This Corollary gives a sufficient condition on a cyclically and null Killing symmetric space-time to be in Kundt's class,[3] moreover it states that in physically important axisymmetric space-times describing *pp* waves[3] the Killing vector K must be orthogonal to the orbits of axisymmetry.

Finally, it is worth noting that the solution of Lukács, Perjés, and Sebestyén[4] has a twisting null Killing vector and the second Killing vector is a cyclic one on an open domain. These vectors commute and have nonzero inner product, as it must be according to our Theorem and its Corollary.

Note added in proof: For the sake of completeness it should be noted that the additional timelike Killing vector $\widetilde{K}$ is independent of K and X on U; i.e., there are not functions α and β on U for which $\widetilde{K} = \alpha K + \beta X$ would hold.

ACKNOWLEDGMENTS

I am indebted to Dr. Á. Sebestyén, Dr. Z. Perjés, Dr. B. Lukács, and Dr. Á. Tóth for useful discussions.

[1]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U.P., Cambridge, 1973).
[2]B. Carter, Commun. Math. Phys. **17**, 233 (1970).
[3]D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (VEB Deutsche Verlag der Wissenschaften, Berlin, 1980).
[4]B. Lukács, Z. Perjés, and Á. Sebestyén, J. Math. Phys. **22**, 1294 (1981).
[5]G. Lessner, Gen. Relativ. Gravit. **18**, 899 (1986).
[6]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, London, 1963), Vol. 1.
[7]N. Hicks, *Notes on Differential Geometry* (Van Nostrand–Reinhold, New York, 1965).

The Bondi metric and conformal motions

A. Di Prisco,[a] L. Herrera,[a] and J. Jimenez
Departamento de Fisica, Facultad de Ciencias, Universidad Central de Venezuela, Caracas, Venezuela
V. Galina
Simon Fraser University, Theoretical Science Institute, Burnaby, British Columbia, Canada V5A 1S6
J. Ibáñez
Departamento de Física, Facultad de Ciencias, Universidad de las Islas Baleares, 07071 Palma de Mallorca, Spain

(Received 23 October 1986; accepted for publication 1 July 1987)

It is shown that there exist no physically significant solutions of the Einstein vacuum field equations, except the trivial one (i.e., Minkowski), for axially and reflection symmetric space-times (Bondi metric) admitting a one-parameter group of conformal motions globally defined on S^2.

I. INTRODUCTION

The problem of finding exact solutions of the Einstein equations describing gravitational radiation from bounded sources is, certainly, one of the most important unsolved problems in relativity.

Due to the complexities created by the nonlinearities of the theory, it is almost imperative to introduce additional (*ad hoc*) assumptions in order to integrate the Einstein equations. Usually restrictions are introduced in the form of different kinds of symmetries (e.g., isometries). However, it is known that the existence of an additional Killing vector causes all the physically significant Bondi metrics to be nonradiative[1–3] (radiative Bondi metrics admitting an additional Killing vector do exist, but present angular singularities). Thus if one is interested in physically significant radiative solutions one has to "reduce" the symmetry of the space-time.

It then seems interesting to consider the existence of a conformal Killing vector field as an additional restriction. This choice is also suggested by the fact that for spherically symmetric space-times with fluids (admitting a one-parameter group of conformal motions) interesting solutions have been obtained (see Ref. 4 and references therein).

However, as we shall show in this paper, the situation for the vacuum Bondi-type metrics is rather disappointing.

In fact, excluding all the singular solutions that do exist, the only completely regular (on the sphere) metric, compatible with a conformal Killing vector, is the Minkowski metric.

II. PRELIMINARY CONSIDERATIONS

Here we briefly present Bondi's formalism and write down the equations for a conformal Killing vector.

A. Bondi's formalism

Let us consider a nonstatic, axially, and reflection symmetric metric[5] which in radiation coordinates takes the form

$$ds^2 = (Vr^{-1}e^{2b} - U^2r^2e^{2g})du^2 + 2e^{2b}\,du\,dr + 2Ur^2e^{2g}\,du\,d\theta - r^2(e^{2g}\,d\theta^2 + e^{-2g}\sin^2\theta\,d\phi^2), \tag{1}$$

where U, V, g, and b are functions of u, θ, and r. Here $u \equiv x^0$ is the timelike coordinate, $r = x^1$ is a null coordinate, and θ and ϕ are two angular coordinates. The condition that the solution be truly isolated requires that the metric functions be regular everywhere; in particular on the polar axis ($\theta = 0,\pi$), which means that V, b, $(U/\sin\theta)$, and $(g/\sin^2\theta)$ are regular functions of $\cos\theta$ as $\cos\theta = \pm 1$. We would like to stress that this condition will be satisfied all through this paper and that its violation would lead to a completely different set of results.

It is well known that the field equations split into two groups: the main equations and the supplementary conditions (actually there is also a trivial equation). The former read

$$b_1 = \tfrac{1}{2}r(g_1)^2, \tag{2}$$

$$[r^4e^{2(g-b)}U_1]_1 - 2r^2[b_{12} - g_{12} + 2g_1g_2 - 2b_2r^{-1} - 2g_1\cot\theta] = 0, \tag{3}$$

$$2V_1 + \tfrac{1}{2}r^4e^{2(g-b)}(U_1)^2 - r^2U_{12} - 4rU_2 - r^2U_1\cot\theta - 4r\cot\theta U + 2e^{2(b-g)}[-1 - (3g_2 - b_2)\cot\theta - g_{22} + b_{22} + (b_2)^2 + 2g_2(g_2 - b_2)] = 0, \tag{4}$$

$$2r(rg)_{01} + (1 - rg_1)V_1 - (rg_{11} + g_1)V - r(1 - rg_1)U_2 - r^2(\cot\theta - g_2)U_1 + r(2g_{12}r + 2g_2 + rg_1\cot\theta - 3\cot\theta)U + e^{2(b-g)}[-1 - (3g_2 - 2b_2)\cot\theta - g_{22} + 2g_2(g_2 - b_2)] = 0, \tag{5}$$

differentiation with respect to u, r, θ, and ϕ is denoted by subscripts 0, 1, 2, and 3, respectively.

Next, if one assumes that the metric functions may be expanded in terms of series in powers of r^{-1}, then one obtains, using (2)–(5),

[a] Postal address: Apartado 80793, Caracas 1080A, Venezuela.

 0022-2488/87/112692-05$02.50

$$g = c(u,\theta)r^{-1} + [C(u,\theta) - \tfrac{1}{6}c^3]r^{-3} + O(r^{-4}), \qquad (6)$$

$$b = -(c^2/4)r^{-2} + O(r^{-4}), \qquad (7)$$

$$U = -(c_2 + 2c\cot\theta)r^{-2} + [2N(u,\theta) + 3cc_2 + 4c^2\cot\theta]r^{-3} + O(r^{-4}), \qquad (8)$$

$$V = r - 2M(u,\theta) - [N_2 + N\cot\theta - (c_2)^2 - 4cc_2\cot\theta - \tfrac{1}{2}c^2(1 + 8\cot\theta)^2]r^{-1} + O(r^{-3}), \qquad (9)$$

with

$$4C_0 = 2c^2c_0 + 2cM + N\cot\theta - N_2, \qquad (10)$$

and the three arbitrary functions of integration M, N, and c are related by the two supplementary conditions,

$$M_0 = -(c_0)^2 + \tfrac{1}{2}(c_{22} + 3c_2\cot\theta - 2c)_0, \qquad (11)$$

$$-3N_0 = M_2 + 3c_{02} + 4cc_0\cot\theta + c_0c_2, \qquad (12)$$

Next, the associated tetrad may be written as

$$l^\mu = e^{-2b}\delta^\mu_1, \quad n^\mu = \delta^\mu_0 - (V/2r)\delta^\mu_1 + U\delta^\mu_2, \quad m^\mu = (1/r\sqrt{2})(e^{-g}\delta^\mu_2 + ie^g\csc\theta\,\delta^\mu_3) \qquad (13)$$

or in covariant components

$$l_\mu = \delta^0_\mu, \quad n_\mu = (Ve^{2b}/2r)\delta^0_\mu + e^{2b}\delta^1_\mu, \quad m_\mu = (r/\sqrt{2})(Ue^g\delta^0_\mu - e^g\delta^2_\mu - ie^{-g}\sin\theta\,\delta^3_\mu). \qquad (14)$$

For the spin coefficients we get[6]

$$\kappa = \epsilon = 0, \quad \rho = -e^{-2b}/r, \quad \sigma = -e^{-2b}g_1,$$

$$\gamma = (e^{-2b}/2)[(Ve^{2b}/2r)_1 - (e^{2b})_0 - U(e^{2b})_2],$$

$$\alpha = \frac{1}{4}\left[\frac{re^g e^{-2b}}{\sqrt{2}}U_1 - \frac{\sqrt{2}e^{-g}}{r}b_2 - \frac{\sqrt{2}e^{-g}}{r}(\cot\theta - g_2)\right],$$

$$\beta = \frac{1}{4}\left[\frac{re^g e^{-2b}}{\sqrt{2}}U_1 - \frac{\sqrt{2}e^{-g}}{r}(b_2 + g_2 - \cot\theta)\right],$$

$$\tau = \frac{e^{-2b}}{2\sqrt{2}}\left[re^gU_1 - \frac{2e^{-g}e^{2b}}{r} - b_2\right],$$

$$\pi = \frac{e^{-2b}}{2\sqrt{2}}\left[re^gU_1 + \frac{2e^{-g}e^{2b}}{r}b_2\right],$$

$$\lambda = U\left[g_2 - \frac{\cot\theta}{2}\right] + g_0 + \frac{1}{2}\left[U_2 - \frac{V}{r}g_1\right],$$

$$\mu = \frac{U}{2}\cot\theta - \frac{V}{2r^2} + \frac{U_2}{2}, \quad \nu = \frac{e^{2b}V_2e^{-g}}{2\sqrt{2}r^2}. \qquad (15)$$

Now, the tetrad components of the Lie derivatives of the metric tensor with respect to a general vector field ξ^α are[7]

$$\tfrac{1}{2}n^\alpha n^\beta \underset{\xi}{\mathcal{L}} g_{\alpha\beta} \equiv 2A\,\mathrm{Re}(\gamma) + 2\,\mathrm{Re}(C\bar\nu) + \Delta A, \qquad (16)$$

$$l^\alpha n^\beta \underset{\xi}{\mathcal{L}} g_{\alpha\beta} \equiv 2A\,\mathrm{Re}(\epsilon) - 2B\,\mathrm{Re}(\gamma) + 2\,\mathrm{Re}(C\bar\pi) - 2\,\mathrm{Re}(C\tau) + DA + \Delta B, \qquad (17)$$

$$-n^\alpha m^\beta \underset{\xi}{\mathcal{L}} g_{\alpha\beta} \equiv B\bar\nu - A(\bar\alpha + \beta + \tau) - C\bar\lambda - \bar C\mu - 2iC\,\mathrm{Im}(\gamma) - \delta A + \Delta\bar C, \qquad (18)$$

$$\tfrac{1}{2}l^\alpha l^\beta \underset{\xi}{\mathcal{L}} g_{\alpha\beta} \equiv -2B\,\mathrm{Re}(\epsilon) - C\kappa - \overline{C\kappa} + DB, \qquad (19)$$

$$-l^\alpha m^\beta \underset{\xi}{\mathcal{L}} g_{\alpha\beta} \equiv -A\kappa + B(\bar\alpha + \beta + \bar\pi) + C\sigma + \bar C[\bar\rho - 2i\,\mathrm{Im}(\epsilon)] - \delta B + D\bar C, \qquad (20)$$

$$m^\alpha \bar m^\beta \underset{\xi}{\mathcal{L}} g_{\alpha\beta} \equiv 2A\,\mathrm{Re}(\rho) - 2B\,\mathrm{Re}(\mu) + 2\,\mathrm{Re}[C(\bar\alpha - \beta)] - \delta C - \overline{\delta C}, \qquad (21)$$

$$\tfrac{1}{2}m^\alpha m^\beta \underset{\xi}{\mathcal{L}} g_{\alpha\beta} \equiv A\sigma - B\bar\lambda - \delta\bar C - \bar C(\bar\alpha - \beta), \qquad (22)$$

with

$$\xi^\alpha = Al^\alpha + Bn^\alpha + Cm^\alpha + \bar C\bar m^\alpha$$

[this C is not to be confused with the function $C(u,\theta)$ in Eqs. (6)–(10)], and

$$D \equiv l^\mu\partial_\mu, \quad \Delta \equiv n^\mu\partial_\mu, \quad \delta \equiv m^\mu\partial_\mu, \quad \bar\delta \equiv \bar m^\mu\partial_\mu.$$

B. The equations for the conformal Killing vector

As was stated in the Introduction, we shall assume that the space-time admits a one-parameter group of conformal motions, i.e.,

$$\underset{\xi}{\mathcal{L}} g_{\alpha\beta} = \psi g_{\alpha\beta}, \qquad (23)$$

where ψ is an arbitrary function of u, θ, and r.

Taking into account (1) and (15)–(22), Eqs. (23) read

$$e^{-2b}B_1 = 0, \qquad (24)$$

$$-Be^{-2b}[(Ve^{2b}/2r)_1 - (e^{2b})_0 - U(e^{2b})_2] + 2\sqrt{2}\,\mathrm{Re}(C)(b_2/r)e^{-g} + e^{-2b}A_1 + B_0 + Ub_2 = \psi, \qquad (25)$$

$$\frac{Br}{\sqrt{2}}U_1e^ge^{-2b} - Cg_1e^{-2b} - \bar C\frac{e^{-2b}}{r} - \frac{\sqrt{2}}{2r}e^{-g}B_2 + e^{-2b}\bar C_1 = 0, \qquad (26)$$

$$Ae^{-2b}\left[\left(\frac{Ve^{2b}}{2r}\right)_1 - (e^{2b})_0 - U(e^{2b})_2\right] + \mathrm{Re}(C)\frac{\sqrt{2}}{2}V_2\frac{e^{-g}e^{2b}}{r^2} + A_0 - \frac{V}{2r}A_1 + UA_2 = 0, \qquad (27)$$

$$-A\left[\frac{r}{\sqrt{2}}U_1e^ge^{-2b} - \frac{\sqrt{2}}{r}b_2e^{-g}\right] + \frac{BV_2e^{-g}e^{2b}}{2\sqrt{2}r^2} - C\left[U\left(g_2 - \frac{\cot\theta}{2}\right) + g_0 + \frac{1}{2}\left(U_2 - \frac{Vg_1}{r}\right)\right] - \bar C\left[\frac{U}{2}\cot\theta + \frac{U_2}{2} - \frac{V}{2r^2}\right] - \frac{1}{\sqrt{2}r}e^{-g}A_2 + \bar C_0 - \frac{V}{2r}\bar C_1 + U\bar C_2 = 0, \qquad (28)$$

$$-Ag_1e^{-2b} - B\left[U\left(g_2 - \frac{\cot\theta}{2}\right) + g_0 + \frac{1}{2}\left(U_2 - \frac{Vg_1}{r}\right)\right] + \bar C\frac{\sqrt{2}}{2r}e^{-g}(\cot\theta - g_2) - \frac{1}{\sqrt{2}r}e^{-g}\bar C_2 - \frac{ie^g\bar C_3}{\sqrt{2}r\sin\theta} = 0, \qquad (29)$$

$$-\frac{2Ae^{-2b}}{r}-B\left(U_2+U\cot\theta-\frac{V}{r^2}\right)$$
$$-\operatorname{Re}(C)\frac{\sqrt{2}}{r}e^{-g}(\cot\theta-g_2)$$
$$-\frac{2}{\sqrt{2}r}\operatorname{Re}\left(e^{-g}C_2+i\frac{e^g}{\sin\theta}C_3\right)=-\psi. \quad (30)$$

Before closing this section it is worth making the following remark.

Since the scale factor ψ is independent of the angle coordinate ϕ (we are dealing with axially symmetric metrics), then taking derivatives of Eqs. (24)–(30) with respect to ϕ, we see that if A, B, and C are functions of ϕ, condition (23) implies the existence of a Killing vector with components

$$\tilde{\xi}^\mu=A_3l^\mu+B_3n^\mu+C_3m^\mu+\overline{C}_3\bar{m}^\mu. \quad (31)$$

However, since the Bondi metric (in its most general form) admits only one Killing vector (associated with the axial symmetry)

$$\tilde{\xi}^\mu=\delta^\mu_\phi=-(i\sqrt{2}/2)re^{-g}\sin\theta(m^\mu-\bar{m}^\mu) \quad (32)$$

we have assumed that, in principle, only C may depend on ϕ.

III. FINDING SOLUTIONS

The procedure to obtain solutions, under restriction (23), is very simple. We shall feed Eqs. (24)–(31) with the series expansion (6)–(9), assuming that the functions A, B, C, and ψ can be written as a series in $1/r$ with both positive and negative powers. Then solving for each order in the expansion we shall find restrictions on the news function c_0, leading to the possible solutions. We shall use superscripts to denote the coefficients of each power,

$$\begin{pmatrix}A\\B\\C\\\psi\end{pmatrix}=\cdots+\begin{pmatrix}A\\B\\C\\\psi\end{pmatrix}^{(-N)}r^N+\cdots+\begin{pmatrix}A\\B\\C\\\psi\end{pmatrix}^{(-1)}r+\begin{pmatrix}A\\B\\C\\\psi\end{pmatrix}^{(0)}+\begin{pmatrix}A\\B\\C\\\psi\end{pmatrix}^{(1)}r^{-1}+\cdots. \quad (33)$$

Further, we shall consider conformal Killing vector fields that asymptotically approach the homothetic vector field, i.e.,

$$\psi=\psi^{(0)}+\psi^{(1)}r^{-1}+\cdots+\psi^{(N)}r^{-N}\quad (N>0).$$

It follows from (27) that $\operatorname{Re}(C)$ does not depend on ϕ, and from the imaginary part of (29)

$$(\operatorname{Im}C)_2/\operatorname{Im}C=\cot\theta-g_2. \quad (34)$$

Next from the imaginary part of (26) and (28) we obtain, respectively,

$$(\operatorname{Im}C)_1/\operatorname{Im}C=1/r-g_1 \quad (35)$$

and

$$-\left[U\left(g_2-\frac{\cot\theta}{2}\right)+g_0+\frac{1}{2}\left(U_2-\frac{Vg_1}{r}\right)\right]\operatorname{Im}(C)$$
$$+\left[\frac{U}{2}\cot\theta+\frac{U_2}{2}-\frac{V}{2r^2}\right]\operatorname{Im}(C)-\operatorname{Im}(C_0)$$
$$+\frac{V}{2r}\operatorname{Im}(C_1)-U\operatorname{Im}(C_2)=0 \quad (36)$$

and, feeding back (34) and (35) into (36),

$$(\operatorname{Im}C)_0/\operatorname{Im}C=-g_0. \quad (37)$$

Solving (34), (35), and (37) and taking into account that $\operatorname{Im}C$ is, at most, linear in ϕ, as it follows from (30), we obtain

$$\operatorname{Im}C=r\sin\theta e^{-g}(a\phi+b),\quad a,b=\text{const}. \quad (38)$$

Next, it follows at once from (24) that

$$B=B(u,\theta). \quad (39)$$

Then, we obtain from the real part of (26), solving for the order $O(r^{-1})$ and $O(r^{-2})$,

$$F^{(-N)}=0,\quad N\geqslant2, \quad (40)$$

$$O(r^{-1}):\ -F^{(0)}+cF^{(-1)}-B_2/\sqrt{2}=0, \quad (41)$$

$$O(r^{-2}):\ -2F^{(1)}+cF^{(0)}+(2B/\sqrt{2})(c_2+2c\cot\theta)$$
$$+B_2c/\sqrt{2}=0, \quad (42)$$

where

$$\operatorname{Re}C=\sum_{-\infty}^{1}F^{(-N)}r^N. \quad (43)$$

We may combine (41) and (42) to obtain

$$2F^{(1)}=c^2F^{(-1)}+(2B/\sqrt{2})(c_2+2c\cot\theta). \quad (44)$$

Next, from the real part of (27) and taking into account (25) and (29),

$$A^{(-N)}=0,\quad \text{for } N\geqslant2,$$
$$O(r):\ A_0^{(-1)}=0, \quad (45)$$
$$O(1):\ A_0^{(0)}-A^{(-1)}/2=0, \quad (46)$$
$$O(r^{-1}):\ A^{(-1)}(2M+cc_0)-\sqrt{2}M_2F^{(-1)}+A_0^{(1)}$$
$$-A_2^{(-1)}(c_2+2c\cot\theta)=0. \quad (47)$$

Working on (25), in the same way, we obtain

$$O(1):\ A^{(-1)}+B_0=\psi^{(0)}, \quad (48)$$
$$O(r^{-1}):\ \psi^{(1)}=0, \quad (49)$$
$$O(r^{-2}):\ -B(M+cc_0)+\sqrt{2}F^{(-1)}(-cc_2)-A^{(1)}$$
$$+A^{(-1)}c^2/2-(c_2+2c\cot\theta)B_2=\psi^{(2)}. \quad (50)$$

Next, from the real part of (28),

$$O(r):\ F_0^{(-1)}=0\Rightarrow F^{(-1)}=\overset{(-1)}{F}(\theta), \quad (51)$$

$$O(1):\ -F^{(-1)}c_0-A_2^{(-1)}/\sqrt{2}+F_0^{(0)}=0, \quad (52)$$

and using (48) and (41) in (52)

$$\psi_2^{(0)} = 0. \tag{53}$$

From the study of Eq. (30) we obtain

$$O(1):\ -2A^{(-1)} - F^{(-1)}\sqrt{2}\cot\theta - (2/\sqrt{2})F_2^{(-1)} + 2a/\sqrt{2} = -\psi^{(0)}, \tag{54}$$

or using (48),

$$2B_0 - \sqrt{2}(F^{(-1)}\cot\theta + F_2^{(-1)}) = \psi^{(0)} + \sqrt{2}a \tag{55}$$

and

$$O(r^{-1}):\ -2A^{(0)} + B - F^{(0)}\sqrt{2}\cot\theta - F^{(-1)}\sqrt{2}(-c_2 - c\cot\theta) - \sqrt{2}F_2^{(0)} + \sqrt{2}cF_2^{(-1)} = -\psi^{(1)}, \tag{56}$$

or using (41),

$$B + \cot\theta B_2 + B_{22} = 2A^{(0)}. \tag{57}$$

Finally, from the real part of (29), we get

$$O(1):\ F^{(-1)}\cot\theta - F_2^{(-1)} = a, \tag{58}$$

$$O(r^{-1}):\ A^{(-1)}c - Bc_0 + F^{(0)}(\sqrt{2}/2)\cot\theta - F^{(-1)}(\sqrt{2}/2)(c_2 + c\cot\theta) - (1/\sqrt{2})F_2^{(0)} + (1/\sqrt{2})cF_2^{(-1)} = 0, \tag{59}$$

or using (41) and (48),

$$-2B_0c - 2Bc_0 + 2\psi^{(0)}c - \cot\theta B_2 - 2\sqrt{2}F^{(-1)}c_2 + B_{22} = 0. \tag{60}$$

We can now integrate Eq. (58) to obtain

$$F^{(-1)} = H\sin\theta - a(\ln|\tan(\theta/2)|)\sin\theta, \tag{61}$$

where $H=$ const.

Feeding back (61) into (55) we obtain, for B_0,

$$2B_0 = 2\sqrt{2}H\cos\theta - 2\sqrt{2}a\ln|\tan(\theta/2)|\cos\theta + \psi^{(0)}. \tag{62}$$

Thus for our conformal Killing vector field to be globally defined in S^2, we must take $a=0$. Next, it follows from (53), (48), (45), and (62) that

$$\psi^{(0)} = \text{const}. \tag{63}$$

Then feeding back (63) into (48) and integrating, we get

$$B(u,\theta) = B_0(\theta)u + B'(\theta), \tag{64}$$

where $B'(\theta)$ is a function of integration that can be eliminated by means of a supertranslation.

We may now write Eq. (60) as

$$c\cos\theta + c_0u\cos\theta + c_2\sin\theta + (1/2\sqrt{2}H)\psi^{(0)}(c_0u - c) = 0, \tag{65}$$

where we have used (61) and (62).

Finally, taking derivatives of (65) with respect to u, we obtain

$$2c_0\cos\theta + c_{00}u\cos\theta + c_{20}\sin\theta = -(c_{00}/2\sqrt{2}H)u\psi^{(0)}. \tag{66}$$

This last equation imposes a severe restriction on the news function c_0. Indeed, for any physically plausible solution c_0 should be different from zero in a finite timelike interval (say $[u_1,u_2]$) and vanish outside this inverval. Thus for some $u=\tilde{u}\in[u_1,u_2]$, c_0 has a maximum and $c_{00}(\tilde{u})=0$.

We can solve (66) for $c_0(\tilde{u},\theta)$ to obtain

$$c_0(\tilde{u},\theta) = I/\sin^2\theta, \tag{67}$$

where $I=$ const.

Thus if we restrict ourselves to bounded sources we must put $I=0$, which means, since c_0 has a maximum at $u=\tilde{u}$, that $c_0(u,\theta)=0$ for any $u\in[u_1,u_2]$. Next feeding back this result into Eq. (65) we get

$$c_2\sin\theta = c(\alpha - \cos\theta) \tag{68}$$

with

$$\alpha \equiv \psi^{(0)}/2\sqrt{2}H,$$

which after integration yields

$$c = K[\tan(\theta/2)]^{\alpha}/\sin\theta, \tag{69}$$

with $K=$ const.

Again, excluding all possible singular solutions we put $K=0$, and we have $c_0=0$. At this point it is worth making the following comment: singular solutions admitting nonvanishing news do exist, for example,

$$c_0 = K/\sin^2\theta, \tag{70}$$

where $K=$ const, or also

$$c_0 = (\sin^K\theta/u^P)(\tan\theta/2)^{P\alpha}, \tag{71}$$

with $K,P=$ const, and

$$P = K+2, \quad \alpha = \psi^{(0)}/2\sqrt{2}H.$$

Next, we shall show that the assumption of the existence of a conformal Killing vector not only excludes nonsingular radiative solutions, but all kinds of regular solutions, except the trivial one (Minkowski).

In fact using (41), (44), (57), (61), (62), (64) and the order $O(r^{-1})$ of the real part of (28), we get

$$HM\sin\theta = 0, \tag{72}$$

where we have taken into account that $c=c_0=0$ and B' is eliminated in (64) by means of supertranslation.

Thus either $M=0$, which leads together with $c=0$, to a flat space-time, or $H=0$. In this case we have from (61)

$$F^{(-1)} = 0 \tag{73}$$

and from the order $O(r^{-2})$ in (30),

$$A^{(1)} = -(\psi^{(0)}M/2)u, \tag{74}$$

or

$$A_0^{(1)} = -(\psi^{(0)}M/2). \tag{75}$$

Then using (46) and (47) together with (75) and (73) we obtain

$$M\psi^{(0)} = 0. \tag{76}$$

Thus we get again the flat space-time, unless our conformal Killing vector becomes a Killing vector field, in which case, obviously, there exist nontrivial solutions.

IV. CONCLUSIONS

We have seen so far that the existence of a one-parameter group of conformal motions appears to be too restrictive a condition if we consider the family of physically significant Bondi metrics.

Once again we should insist on the fact that solutions presenting angular singularities (unbounded sources) and/or metrics corresponding to systems radiating during an infinite period of time do, in principle, exist.

Recently, using completely different approaches, results similar to the one presented in this paper have been reported by Eardley *et al.*[8] and by Garfinkle.[9]

In Ref. 8 it is shown that the only solution of the Einstein vacuum equations, asymptotically flat (spatially) and admitting a conformal Killing field, is the Minkowski space-time. The proof relies on the positive energy theorem and the authors use the $3 + 1$ split.

In Ref. 9, a similar result is obtained assuming the space-time to be asymptotically flat at null infinity and the Bondi energy to be positive.

In this paper we have used the leading terms in the expansion of the Bondi metric to show that the existence of a conformal Killing vector field implies that the space-time is either flat (Minkowski) or nonglobally regular at null infinity.

[1]R. Berezdivin and L. Herrera, Phys. Rev. D **11**, 2063 (1975). It is interesting to note that despite the errors contained in the Killing equations, the main conclusion emerging from this reference is still valid, provided we restrict ourselves to the class of physically significant metrics related to bounded sources.

[2]D. Hobill, J. Math. Phys. **25**, 3527 (1984).

[3]J. Bicak and B. G. Schmidt, J. Math. Phys. **25**, 600 (1984).

[4]L. Herrera and J. Ponce de Leon, J. Math. Phys. **26**, 2018 (1985).

[5]H. Bondi, M. G. J. Van der Burg, and A. W. K. Metzner, Proc. R. Soc. London Ser. A **269**, 21 (1962).

[6]E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

[7]L. Herrera and J. Jimenez, Phys. Rev. D **28**, 2987 (1983).

[8]D. Eardley, J. Isenberg, J. Mardsen, and V. Moncrief, Commun. Math. Phys. **106**, 137 (1986).

[9]D. Garfinkle, J. Math. Phys. **28**, 28 (1987).

A class of conformally flat solutions for a charged sphere in general relativity

Wang Xingxiang
Department of Physics, Anhui Normal University, Wuhu, Anhui, People's Republic of China

(Received 5 February 1987; accepted for publication 3 June 1987)

In this paper a class of conformally flat interior solutions of Einstein–Maxwell equations for a charged sphere is obtained. These solutions satisfy physical conditions inside the sphere.

I. INTRODUCTION

The problem of finding the exact solutions of coupled Einstein–Maxwell equations for the static distribution of a charged sphere has attracted wide attention. These distributions constitute possible sources for a Reissner–Nordström metric that uniquely describe the exterior field of a spherically symmetric charged distribution of matter. Recently some conformally flat solutions of Einstein–Maxwell equations for a charged sphere were presented by Banerjee and Santos[1] and Shi-Chang,[2] but none of these solutions is free from singularity as well as satisfies the energy conditions.

In this paper, a class of conformally flat solutions for a static charged sphere is presented by specifying matter distribution. These metrics are free from singularity and can be matched to the Reissner–Nordström metric. The energy-momentum tensor satisfies the energy conditions, so that the solutions are physically reasonable.

II. FIELD EQUATIONS AND SOLUTIONS

We will use, as usual, the standard coordinate, and the line element is given by

$$ds^2 = e^\nu dt^2 - e^\lambda dr^2 - r^2 d\vartheta^2 - r^2 \sin^2\vartheta\, d\varphi^2. \tag{1}$$

Here, ν and λ are functions of r alone. The resulting Einstein–Maxwell equations are[3]

$$8\pi\rho + E^2 = e^{-\lambda}\left(\frac{\lambda'}{r} - \frac{1}{r^2}\right) + \frac{1}{r^2}, \tag{2}$$

$$8\pi p - E^2 = e^{-\lambda}\left(\frac{\nu'}{r} + \frac{1}{r^2}\right) - \frac{1}{r^2}, \tag{3}$$

$$8\pi p + E^2 = e^{-\lambda}\left(\frac{\nu''}{2} + \frac{\nu'^2}{4} - \frac{\lambda'\nu'}{4} + \frac{\nu' - \lambda'}{2r}\right), \tag{4}$$

and

$$Er^2 = q(r) = 4\pi\int_0^r \sigma r^2 e^{\lambda/2}\, dr, \tag{5}$$

where p is the interior presure, ρ is the mass density, σ is the charge density, and $q(r)$ represents the total charge contained within a sphere of radius r.

Since it is conformally flat, the vanishing of the Weyl tensor gives[2]

$$\frac{e^\lambda}{r^2} - \frac{1}{r^2} - \frac{\nu'^2}{4} + \frac{\nu'\lambda'}{4} - \frac{\nu''}{2} - \frac{1}{2r}(\lambda' - \nu') = 0. \tag{6}$$

Then subtracting (3) from (4) we have

$$2E^2 = e^{-\lambda}\left[\frac{\nu''}{2} - \frac{\lambda'\nu'}{4} + \frac{\nu'^2}{4} - \frac{1}{r^2} - \frac{1}{2r}(\lambda' + \nu') + \frac{e^\lambda}{r^2}\right]. \tag{7}$$

Multiplying Eq. (6) by $e^{-\lambda}$, then adding the result to (7), we get

$$E^2 = \frac{1}{r^2} - e^{-\lambda}\left(\frac{\lambda'}{2r} + \frac{1}{r^2}\right). \tag{8}$$

We have five equations (2)–(5) and (8) and six variables $(\rho, \sigma, p, E^2, \nu, \lambda)$. Hence we have only one free variable. By solving these equations we can obtain the other five, if any variables have been predetermined. In this paper we assume that the mass density is

$$8\pi\rho = 6nC(1 - Cr^2)/[1 + (n-1)Cr^2]^3, \tag{9}$$

where C is a positive constant and n is a parameter ($n = 1,2,3,\ldots$). For each integral value of n, we have a different model. Due to the requirement that $\rho \geqslant 0$, we have a restriction on C, that is $Ca^2 \leqslant 1$, where a is the radius of the sphere. Clearly, the mass density is a monotonically decreasing function of r.

In what follows, defining

$$e^{-\lambda} = Z, \quad e^{\nu/2} = y, \quad Cr^2 = x, \tag{10}$$

Eqs. (2)–(4), (8), and (9) can now be expressed as

$$\frac{8\pi\rho + E^2}{C} = \frac{1-Z}{x} - 2\frac{dZ}{dx}, \tag{11}$$

$$\frac{8\pi p - E^2}{C} = \frac{4Z}{y}\frac{dy}{dx} - \frac{1-Z}{x}, \tag{12}$$

$$\frac{8\pi p + E^2}{C} = \frac{4xZ}{y}\frac{d^2y}{dx^2} + \frac{4Z}{y}\frac{dy}{dx} + \frac{2x}{y}\frac{dZ}{dx}\frac{dy}{dx} + \frac{dZ}{dx}, \tag{13}$$

$$\frac{E^2}{C} = \frac{1-Z}{x} + \frac{dZ}{dx}, \tag{14}$$

$$\frac{8\pi\rho}{C} = \frac{6n(1-x)}{[1+(n-1)x]^3}. \tag{15}$$

From (11) and (14) we have

$$\frac{dZ}{dx} = -\frac{8\pi\rho}{3C}. \tag{16}$$

Substituting (15) into (16), we get

$$e^{-\lambda} = Z = (1-x)^2/[1+(n-1)x]^2. \tag{17}$$

We have taken the constant of integration equal to 0 so that $e^{-\lambda} = 1$ when $r = 0$.

From (12)–(14), we have

$$4x^2 Z\frac{d^2y}{dx^2} + 2x^2\frac{dZ}{dx}\frac{dy}{dx} - \left[x\frac{dZ}{dx} + (1-Z)\right]y = 0. \tag{18}$$

Substitution of (17) into Eq. (18) gives a particular solution of y as

 0022-2488/87/112697-03$02.50

$$y_1 = (1-x)^{n/2}. \tag{19}$$

Then another particular solution of Eq. (18) can be obtained as

$$y_2 = y_1 \int y_1^{-2} Z^{-1/2}\, dx = x(1-x)^{-n/2}. \tag{20}$$

Then the general solution of Eq. (18) is

$$y = B\,[(1-x)^{n/2} + Ax(1-x)^{-n/2}], \tag{21a}$$

also

$$e^{\nu} = B^2[(1-x)^{n/2} + Ax(1-x)^{-n/2}]^2, \tag{21b}$$

where A and B are constants. From (12), (14), and (21) the pressure, electric field intensity, and charge density can be obtained as

$$\frac{8\pi p}{C} = \frac{2(1-x)\{2(1-x)[1+(n-1)x] + n(n-1)x^2\}}{[1+(n-1)x]^3[(1-x)^n + Ax]}\left\{A - \frac{n[2+(n-1)x](1-x)^n}{2(1-x)[1+(n-1)x] + n(n-1)x^2}\right\}, \tag{22}$$

$$E^2 = Cnx[(3n-2) + (n-1)(n-2)x]/[1+(n-1)x]^3, \tag{23}$$

$$\sigma = \pm \frac{C\sqrt{n}(1-x)[3(3n-2) + 4(n-1)(n-2)x + (n-1)^2(n-2)x^2]}{4\pi[1+(n-1)x]^{7/2}[(3n-2)+(n-1)(n-2)x]^{1/2}}. \tag{24}$$

The gravitational field in the exterior region of the static charged fluid sphere of radius a is uniquely described by the Reissner–Nordström metric

$$ds^2 = \left(1 - \frac{2m}{r} + \frac{Q^2}{r^2}\right)dt^2 - \left(1 - \frac{2m}{r} + \frac{Q^2}{r^2}\right)^{-1} dr^2 - r^2\, d\vartheta^2 - r^2 \sin^2\vartheta\, d\varphi^2. \tag{25}$$

Accordingly, metrics (17) and (22) will describe the field in the interior of a charged fluid sphere of radius a, if and only if the boundary conditions

$$e^{\nu(a)} = e^{-\lambda(a)} = (1 - 2m/a + Q^2/a^2) \tag{26}$$

and

$$p(r=a) = 0 \tag{27}$$

are satisfied across the boundary $r = a$.

From (5) and (23) we have the following expression for total charge:

$$\frac{Q^2}{a^2} = \frac{nx_1^2[(3n-2) + (n-1)(n-2)x_1]}{[1+(n-1)x_1]^3}, \tag{28}$$

where $x_1 = Ca^2$. The constants A and B and total mass m can be determined by considering the boundary conditions (26) and (27), i.e.,

$$A = \frac{n(1-x_1)^n[2+(n-1)x_1]}{2(1-x_1)[1+(n-1)x_1] + n(n-1)x_1^2}, \tag{29}$$

$$B^2 = \frac{(1-x_1)^{n+2}}{[(1-x_1)^n + Ax_1]^2[1+(n-1)x_1]^2}, \tag{30}$$

$$\frac{2m}{a} = 1 + \frac{nx_1^2[(3n-2)+(n-1)(n-2)x_1]}{[1+(n-1)x_1]^3} - \frac{(1-x_1)^2}{[1+(n-1)x_1]^2}. \tag{31}$$

It is clear that the metrics are free from singularity for all values of n inside the sphere. A physically reasonable energy-momentum tensor has to obey the energy conditions

$$-\rho \leqslant p \leqslant \rho, \quad \rho \geqslant 3p. \tag{32}$$

From (15), (22), and (29) we can see that the requirements (32) are satisfied, at least in the case $x_1 \ll 1$. More details of the solutions will be studied for several particular values of n in the next section.

III. SOLUTIONS FOR DIFFERENT VALUES OF n

By taking different values of n, we obtained a very large number of exact solutions of Einstein–Maxwell equations. Here, we write the exact solutions for $n = 1$ and $n = 2$ only, and discuss the physical relevance of these solutions.

A. $n=1$

From (28)–(31) we have

$$A = 1, \quad B^2 = (1-x_1)^3, \tag{33}$$

$$m/a = x_1, \quad Q^2/a^2 = x_1^2 \tag{34}$$

(where $x_1 = Ca^2$). Then the solutions are given by

$$\rho = (3C/4\pi)(1-x), \quad \sigma = \pm\rho, \tag{35}$$

$$E^2 = Cx, \quad p = 0, \tag{36}$$

$$e^{-\lambda} = (1-x)^2, \quad e^{\nu} = (1-x_1)^3/(1-x). \tag{37}$$

It is clear that the solution describes a dust sphere ($p=0$) and proper charge density $\sigma e^{\lambda/2}$ is constant. Also, the De-Raychaudhari requirement $\sigma = \pm\rho$ is fulfilled. This charged dust sphere solution belongs to the class of solutions previously found by Bonnor.[4]

B. $n=2$

The solutions can be written as

$$\rho = \frac{3C}{2\pi}\frac{(1-x)}{(1+x)^3}, \quad \sigma = \pm\frac{3\sqrt{2}C}{2\pi}\frac{(1-x)}{(1+x)^{7/2}}, \tag{38}$$

$$E^2 = 8Cx/(1+x)^3, \tag{39}$$

$$\frac{8\pi p}{C} = \frac{4(1-x)}{(1+x)^3[(1-x)^2 + Ax]} \times [A - (2+x)(1-x)^2], \tag{40}$$

$$e^{-\lambda} = (1-x)^2/(1+x)^2, \quad e^{\nu} = B^2[(1-x) + Ax(1-x)^{-1}]^2, \tag{41}$$

and

$$\frac{Q^2}{a^2} = \frac{8x_1^2}{(1+x_1)^3}, \quad \frac{m}{a} = \frac{2x_1(1+3x_1)}{(1+x_1)^3}, \tag{42}$$

$$A = (2+x_1)(1-x_1)^2, \quad B^2 = \frac{1}{(1+x_1)^6}. \tag{43}$$

This is a charged fluid sphere solution. The metric is free from singluarity and the enegy conditions (32) are satisfied, at least in the case $x_1 < \frac{1}{3}$, i.e., $m/a < \frac{9}{16}$.

[1]A. Banerjee and N. O. Satos, J. Math. Phys. **22**, 824 (1981).
[2]Z. Shi-Chang, Gen. Relativ. Gravit. **15**, 293 (1983).
[3]R. Tikekar, J. Math. Phys. **25**, 1481 (1984).
[4]W. B. Bonnor, Mon. Not. R. Astron. Soc. **129**, 443 (1965).

Relativistic fluids with shear and timelike conformal collineations

K. L. Duggal
Department of Mathematics and Statistics, University of Windsor, Windsor, Ontario, Canada N9B 3P4

(Received 17 March 1987; accepted for publication 1 July 1987)

This paper deals with perfect fluid matter plus nonsingular aligned electromagnetic fields admitting a timelike conformal collineation. It is proved that nonvanishing shear is itself a symmetry condition that partially underlines timelike conformal collineations with symmetry vector parallel to the fluid velocity vector. The results provide a natural counterpart to the study on shear-free space-times. It is also shown that the present study is applicable to a variety of physical problems.

I. INTRODUCTION

A space-time (V_4,g_{ab}) admits a one-parameter group of conformal motions (Conf M) generated by a vector field ξ if

$$\mathscr{L}_\xi g_{ab} = 2\Psi g_{ab} \quad (a,b=0,1,2,3), \tag{1}$$

where $\mathscr{L}$ is the Lie-derivative operator and Ψ is an arbitrary function of the coordinates. Since the celebrated work of Weyl,[1] in 1921, conformal symmetry property (preserving angles but not scale lengths at different points) has been an essential geometric prescription for a good part of physics. For example, all equations of massless particles,[2] including the Yang–Mills equations,[3] are conformally invariant. In the context of general relativity, conformal symmetry has been very useful in finding exact solutions[4–7] of field equations. In particular, for the study of fluid space-times, conformal symmetry plays the role of preserving the continuity of the matter flow at critical points of transition during a change of state. A simple example is of a liquid which changes to gas when heated through its boiling point. When the pressure is raised, the transition becomes less and less abrupt, until at a critical pressure it is continuous. But, at this critical point, the density fluctuations occur at all length scales. It is a remarkable phenomenon of universality that most physical systems do respond in a natural way to the local conformal symmetry (scale transformations) at those critical points. Thus the conformal symmetry measures the response of the fluid subject to large density fluctuations and describes the leading finite size correction to scaling at critical points. We also mention Cahill and Taub[8] who pointed out that spherical symmetric perfect fluid solutions admitting a homothetic (special case of conformal) symmetry represents the relativistic generalization of the self-similar solutions of classical hydrodynamics (cf. also Eardley[9] and Wainwright[10]).

It is well-known that every (Conf M) must satisfy

$$\mathscr{L}_\xi \{ {}_b{}^a{}_c \} = \delta^a{}_b \Psi_{;c} + \delta^a{}_c \Psi_{;b} - g_{bc} g^{ad} \Psi_{;d}, \tag{2}$$

but the converse is not necessarily true. However, condition (2) is equivalent to

$$\mathscr{L}_\xi g_{ab} = 2\Psi g_{ab} + h_{ab}, \quad h_{[ab]} = 0 = h_{ab;c}, \tag{3}$$

where h_{ab} is a symmetric, parallel (and therefore Killing) tensor associated with ξ.

Definition: A space-time V_4 admits a conformal collineation (Conf C) generated by an affine conformal vector (ACV) if (2) or equivalently (3) holds.

Thus, in general, conformal symmetry extends to conformal collineation (which may not pull back Conf M) as a result of the action of $\mathscr{L}_\xi$ on the connection coefficients $\{ {}_b{}^a{}_c \}$. Using this information and a paper by Herrera *et al.*,[4] Duggal and Sharma[11] indicated the possibility of new anisotropic and isotropic solutions if (Conf M) is replaced by (Conf C). Motivated by the above, we continue our study on this subject. The objective is to investigate some properties of fluids with nonvanishing shear (the results are valid for space-times of general relativity) directly related to (Conf C) with symmetry vector parallel to the fluid velocity vector. We consider a perfect fluid matter with electromagnetic field. This choice is motivated by the following known results in the related study on fluids. Tupper[12] has shown the equivalence of perfect fluids plus electromagnetic fields with viscous fluids under some geometric restrictions. This work was further extended by Maiti and Das[13] to include heat flux. It is also important to note that a particular case of an anisotropic fluid is that of a viscous fluid, characterized by an anisotropic pressure tensor proportional to the shear tensor of the velocity field.[7]

The main emphasis, in this paper, is on the important role of conformal collineations in the study on *fluids with shear* as a counterpart of the *shear-free fluids* characterized by conformal motions. We also show that the present study is applicable to a variety of physical problems in relativistic fluids and astrophysics.

II. ENERGY-MOMENTUM TENSOR

Let the energy momentum tensor T_{ab} be of the form

$$\begin{aligned} T_{ab} &= (\mu + p) u_a u_b + p g_{ab} + E_{ab}, \\ E_{ab} &= F_{ac} F^c{}_b - \tfrac{1}{4} F_{cd} F^{cd} g_{ab}, \end{aligned} \tag{4a}$$

where u^a, μ, p, and F_{ab} are the four-velocity vector, density, pressure, and electromagnetic field. The Maxwell's equations are

$$F_{[ab;c]} = 0 \quad \text{and} \quad F^{ab}{}_{;a} = J^b, \tag{4b}$$

where J^b is the current four-vector. To state, precisely, the type of F_{ab} we recall that at each point of a space-time one can introduce Newman–Penrose (NP) tetrad $\{m^a, \overline{m}^a, l^a, k^a\}$ such that m^a and $\overline{m}^a$ are conjugate complex and l^a, k^a (principal null directions: see Fig. 1) are real null vectors. Here $m^a \overline{m}_a = 1$, $k^a l_a = -1$ (other products zero). Thus $g_{ab} = 2m_{(a} \overline{m}_{b)} - 2k_{(a} l_{b)}$ with Maxwell sca-

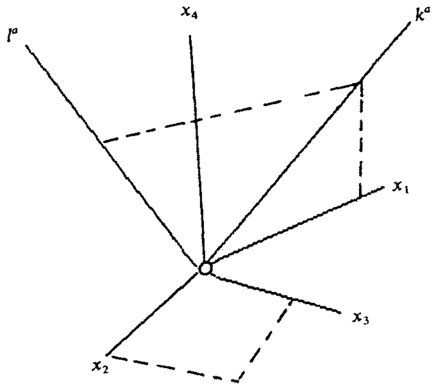

FIG. 1. Principal null directions in the NP tetrad.

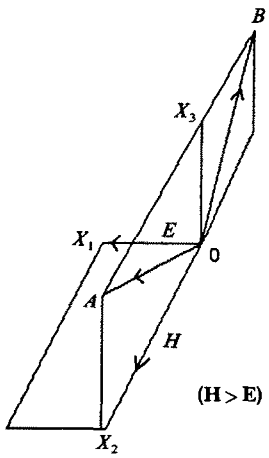

FIG. 3. Representation (OH,OB) in the observer's space of the principal null directions.

lars $\phi_0 = F_{ab} k^a m^b$, $\phi_1 = \frac{1}{2} F_{ab} (k^a l^b + m^a \overline{m}^b)$, $\phi_2 = F_{ab} \times \overline{m}^a l^b$.

In this paper, we assume that F_{ab} is non-null and it corresponds to the family of an aligned ($\phi_0 = 0$), nonradiative ($\phi_2 = 0$) and real or imaginary ($\phi_1 \neq 0$) field. Set $\phi_1 = \phi$. Thus

$$F_{ab} = 2\phi(l_a k_b - k_a l_b) \quad \text{or} \quad F_{ab} = 2\phi(m_a \overline{m}_b - \overline{m}_a m_b) \tag{4c}$$

according as ϕ is real or imaginary. For a flat space, $\mathbf{E}\cdot\mathbf{H} = 0$, $|\mathbf{E}|^2 - |\mathbf{H}|^2 \neq 0$, where E and H are electric and magnetic vectors. Following Synge,[14] we state the following theorem.

Theorem: In a field that satisfies (4) with ϕ real, to a general observer, $\mathbf{E} \perp \mathbf{H}$, $\mathbf{E} > \mathbf{H}$, and $\mathbf{E}$ constitutes an electromagnetic wrench consisting of an electric vector only (see Fig. 2). For imaginary ϕ, $\mathbf{H} > \mathbf{E}$ and the electromagnetic wrench consists of a magnetic vector only (see Fig. 3).

In terms of a local orthonormal frame (u,n,y,z), we have

$$\begin{aligned} &\sqrt{2}u^a = l^a + k^a, \quad \sqrt{2}n^a = k^a - l^a, \\ &\sqrt{2}y^a = m^a + \overline{m}^a, \quad \sqrt{2}z^a = i(\overline{m}^a - m^a), \\ &\sqrt{2}u_a = l_a + k_a, \quad \sqrt{2}n_a = k_a - l_a, \\ &\sqrt{2}y_a = m_a + \overline{m}_a, \quad \sqrt{2}iz_a = m_a - \overline{m}_a, \end{aligned} \tag{4d}$$

where $u^a u_a = -1$ is the four-velocity vector of the fluid and $n^a n_a = y^a y_a = z^a z_a = 1$.

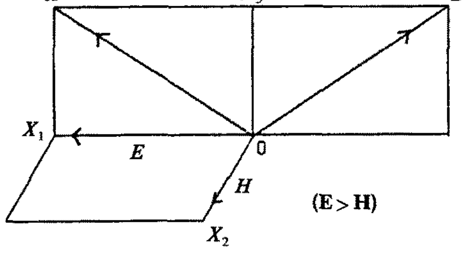

FIG. 2. Representation (OA,OB) in the observer's space of the principal null directions.

III. AN EXAMPLE OF CONFORMAL COLLINEATIONS

We use the following terminology and notations from differential geometry.[15] If f: $M \to N$ is a map of a manifold M into another manifold N (all maps and manifolds are smooth), then Tf: $TM \to TN$ is the tangent map of the corresponding tangent bundles. If g is a covariant tensor field on N, then $f^* g$ denotes its pullback on M. If ξ is a vector field and η is a p-form on M, then $\xi \rfloor \eta$ is the $(p-1)$-form on M obtained by contracting η with ξ. This means that, for $x \in M$ and $V_2,\ldots,V_p \in T_x(M)$, we have

$$(\xi \rfloor \eta)(V_2,\ldots,V_p) = \eta(\xi(x),V_2,\ldots,V_p).$$

For example, if U and V are vector fields on M then $g(U,V) = U \rfloor v$ where $v_a = g_{ab} V^b$ is a one-form. A vector field ξ on M generates a one-parameter group of local transformations, called a flow (f_t). For any sufficiently small $t \in R$, f_t is a diffeomorphism of an open submanifold of M onto another such submanifold. For small t and s, $f_t f_s = f_{t+s}$. Thus for any tensor field Ω we get

$$f_t^*(f_s^* \Omega) = f_{t+s}^* \Omega.$$

Differentiating both sides with respect to s at $s = 0$, we get

$$\frac{d}{dt}(f_t^* \Omega) = f_t^* L_\xi \Omega, \quad \text{where } \mathscr{L}_\xi \Omega = \frac{d}{dt}(f_t^* \Omega)\bigg|_{t=0}.$$

It is well-known that for a p-form η, we have

$$\mathscr{L}_\xi \eta = \xi \rfloor d\eta + d(\xi \rfloor \eta),$$

where d denotes the exterior product. Let M be a space-time V_4. Based on the NP formalism, one can generate a two-dimensional distribution D: $x \to D_x$ of all the principal directions which cut the null cone by the two principal null directions l^a and k^a as indicated in Fig. 4.

As D will contain a timelike vector, D is non-null (see Ref. 16). We assume that V_4 has a conformal structure which is needed to preserve the causal character of D. A conformal structure, along with D, defines another two-dimensional distribution $D^\perp$. Here D and $D^\perp$ define vector bundles over V_4 such that

$$TV_4 = D \otimes D^\perp, \quad D \cap D^\perp = \{0\}.$$

The vector bundle $D^\perp \to V_4$ has also a conformal structure induced from D. Let the flow (f_t), generated by ξ, preserve the conformal structure of $D^\perp$. This means that there exists a

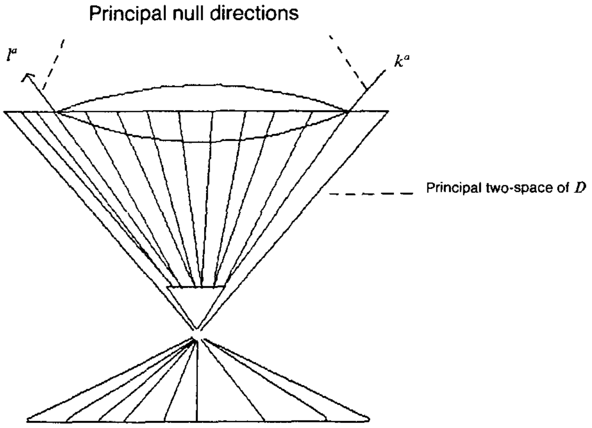

FIG. 4. Representation of the distribution D of all the principal directions cutting the null cone.

positive $h \in R$, such that for any $U,V \in D^{\perp}(x)$, $(f_t g)(U,V) = hg(U,V)$. This is equivalent to the existence of a function h_t and two one-forms ω_t and η_t in D such that $f_t g = h_t g + \omega_t \otimes \eta_t + \eta_t \otimes \omega_t$. By differentiating, this implies

$$\mathscr{L}_\xi g = 2\Psi g + \omega \otimes \eta + \eta \otimes \omega, \tag{5a}$$

where Ψ is a function ω and η are one-forms in D. Let ξ be non-null and generates the flow (f_t) in D. Then, for an orthonormal frame $\{u,n,y,z\}$, η_a and ω_a are in the local two-space generated by $\{u,n\}$ as D cuts the null cone (see Fig. 4). This means that (for some scalars A, B, and C)

$$\mathscr{L}_\xi g_{ab} = 2\Psi g_{ab} + h_{ab}, \tag{5b}$$

$$h_{ab} = A u_a u_b + B n_a n_b + C[u_a n_b + u_b n_a]. \tag{5c}$$

Now we deal with the case when ξ is null. As ξ cannot be complex, it must be along one of the null principal directions and, therefore, belongs to D. So we choose $\xi^a = \propto k^a$. For this case, the local two-spaces of D will be generated by the set $\{k,l\}$ and we consider a pseudo-orthonormal basis $\{k,l,y,z\}$. For details on this, we refer to Ref. 17. Thus (5b) will hold with

$$h_{ab} = A[l_a l_b + k_a k_b] + B[l_a k_b + k_a l_b]. \tag{5d}$$

Conclusion 1: Based on the above discussion, it is possible to generate an ACV ξ^a so that (5b) holds with h_{ab} prescribed by (5c) for non-null and (5d) for null ξ^a provided the NP tetrad is chosen in such a way that $h_{ab;c} = 0$.

Note 1: Most physical space-times inherit spherical or cylindrical or plane symmetry, etc. Thus it is quite reasonable to construct an (ACV) that belongs to two-dimensional distribution D. This will include the most important special cases; namely, timelike, spacelike, and null congruences. For example it is known that Raychaudhuri's equation[18] shows directly the influence of the fluid matter on the convergence of timelike geodesics. This convergence property for the timelike geodesics, together with the related property for null geodesics, is now the root of the singularity theorems[17]; cf. also Ref. 19.

IV. TIMELIKE CONFORMAL COLLINEATIONS

Let $\xi^a = \xi u^a$, $u^a u_a = -1$, and $\xi = \sqrt{(\xi^a \xi_a)} > 0$. Thus ξ^a is a timelike ACV iff (3) holds. Equation (3) may be rewritten equivalently as

$$\xi_{(a;b)} = 2\Psi g_{ab} + h_{ab},\ h_{[ab]} = 0 = h_{ab;c}. \tag{6}$$

We now initiate a comparative study on the kinematic properties of space-times with (Conf M) and (Conf C) followed by indicating the advantages of using (Conf C). For this purpose, we first state a known result on (Conf M) followed by a corresponding result on (Conf C). We prefer to present general results for a timelike curve congruence with unit tangent vector u^a. The results are also valid for fluid space-times for which u^a is its velocity vector. The following result on (Conf M) is known.

Theorem 1 (Oliver and Davis, Ref. 20): A space-time admits a timelike (Conf M) with symmetry vector $\xi^a = \xi u^a$ $(\xi > 0)$ iff

$$\text{(i) } \sigma_{ab} = 0, \quad \text{(ii) } \dot{u}_a = (\log \xi)_{,a} + (\theta/3) u_a, \tag{7}$$

where $\Psi = \xi\theta/3$ and σ_{ab}, θ, and $\dot{u}_a$ are, respectively, the shear, expansion, and the acceleration vector of the timelike congruence generated by u^a.

Let $P_{ab} = g_{ab} + u_a u_b$ be the projection tensor. The following fundamental equation is well known:

$$u_{a;b} = \sigma_{ab} + \tfrac{1}{3}\theta P_{ab} + \omega_{ab} - \dot{u}_a u_b, \tag{8}$$

where ω_{ab} is the vorticity tensor. For details see Ref. 21. For a (Conf C) we have the following corresponding result.

Theorem 2: A space-time admits a timelike (Conf C) with symmetry vector $\xi^a = \xi u^a$ $(u^a u_a = -1, \xi > 0)$ iff

$$\text{(i) } \sigma_{cd} = (2\xi)^{-1}[P^a{}_c p^b{}_d h_{ab} - \tfrac{2}{3}\tilde{\theta} p_{cd}], \quad \text{(ii) } \dot{u}_a = \xi^{-1}[\xi_{,a} + (\dot{\xi}) u_a + h_{bc} u^c P^b{}_a], \tag{9}$$

where $2\tilde{\theta} = h^a{}_a + h_{ab} u^a u^b$ and $\Psi = \tfrac{1}{3}(\xi\theta - \tilde{\theta})$.

Proof: Assume that (6) holds for a timelike symmetry vector $\xi^a = \xi u^a$. Contracting (6) in turn with g^{ab} and $u^a u^b$ gives $\Psi = \tfrac{1}{4}[\xi\theta + \dot{\xi} - \tfrac{1}{2} h^a{}_a]$ and $\Psi = \dot{\xi} + \tfrac{1}{2} h_{ab} u^a u^b$. This implies that $\Psi = \tfrac{1}{3}(\xi\theta - \tilde{\theta})$, where $2\tilde{\theta} = h^a{}_a + h_{ab} u^a u^b$. Now (i) and (ii) follow by contracting (6) with $P^a{}_c P^b{}_d$ and then with $u^a P^b{}_c$ [where we make use of Eq. (8) and the value of Ψ]. Conversely, assuming (9) holds for some covariant constant and symmetric tensor h_{ab} and $\Psi = \tfrac{1}{3}(\xi\theta - \tilde{\theta})$, one can show that ξ^a is an ACV with $\xi^a = \xi u^a$.

Conclusion 2: It is clear from (i) in (9) that there is an advantage of using (Conf C) over (Conf M), in particular reference to the study of space-times with nonvanishing shear.

For an application of Theorem 2 to fluid space-times, we first prove the following for which h_{ab} satisfies (5b). This means

$$h_{ab} = A u_a u_b + B n_a n_b + C(u_a n_b + n_a u_b).$$

Theorem 3: A fluid space-time admits a timelike (Conf C) with symmetry vector $\xi^a = \xi u^a$ and h_{ab} satisfying (5b) iff

$$\text{(i) } \sigma_{ab} = (\sqrt{3}/2)\sigma(n_a n_b - \tfrac{1}{3} P_{ab}),$$

$$\text{(ii) } \dot{u}_a = (\log \xi)_{,a} + (\log \xi) u_a,\ \Psi = \xi(\theta - (\sqrt{3}/2)\sigma). \tag{10}$$

Proof: For h_{ab} specified by (5b), the following will hold: $P^c{}_a P^d{}_b = B n_a n_b$, $2\tilde{\theta} = B$. Now contracting (i) in (9) with σ^{cd} and using the above results, we obtain $B = \sqrt{3}\xi\sigma$, where σ is the magnitude of σ_{ab} defined by $\sigma^2 = \frac{1}{2}\sigma^{cd}\sigma_{cd} \geqslant 0$. Also $h_{bc}u^c P^b{}_a = 0$. Using all these results in (9) we obtain (10).

For further analysis of the above result, we consider the following general form for the energy-momentum tensor of a fluid[22]:

$$T_{ab} = \bar{\mu} u_a u_b + \bar{p} P_{ab} + \Pi_{ab} + q_a u_b + q_b u_a, \qquad (11a)$$

where $\bar{\mu}, \bar{p}, q^a$, and Π_{ab} are the density, the thermodynamic pressure, the energy flux vector, and the anisotropic pressure tensor, respectively ($q_a u^a = 0$, $\Pi_{ab} u^b = 0$, $\Pi^a{}_a = 0$).

A particular case of (11a) is viscous fluids,[7] such that

$$\Pi_{ab} = -2\eta\sigma_{ab}, \qquad (11b)$$

where ($\eta > 0$) is the kinematical viscosity. The expression (11a) along with (11b) establishes the relativistic equivalent to the Navier–Stokes theory of Newtonian fluid mechanics.

Another special case is a preferred direction of the anisotropic pressure with no energy flux ($q_a = 0$), for which[4,11]

$$T_{ab} = \bar{\mu}\, u_a u_b + p_{\parallel} n_a n_b + p_{\perp} S_{ab}, \qquad (11c)$$

where n^a is along the preferred direction, $S_{ab} = P_{ab} - n_a n_b$ is the projection tensor into the two-plane of the pressure isotropy ($S_{ab}u^b = 0 = S_{ab}n^b$), and $P_{\parallel}$ and $P_{\perp}$ are the pressure along and orthogonal to n^a, respectively. The energy-momentum tensor (11a) reduces to (11c) when

$$\text{(i) } \bar{p} = \tfrac{1}{3}(p_{\parallel} + 2p_{\perp}), \qquad \text{(ii) } \Pi_{ab} = (p_{\perp} - p_{\parallel})\left[\tfrac{1}{3}P_{ab} - n_a n_b\right]. \qquad (11d)$$

In this paper, we consider the following energy-momentum tensor:

$$T_{ab} = \mu u_a u_b + p P_{ab} + E_{ab},$$
$$E_{ab} = \pm\phi^2(u_a u_b - n_a n_b + S_{ab}), \qquad (12)$$
$$S_{ab} = y_a y_b + z_a z_b, \quad P_{ab} = S_{ab} + n_a n_b,$$

where μ and p are the energy density, and isotropic pressure, and + or − represent the real or imaginary Maxwell scalar ϕ. For more details see Sec. II. The energy-momentum tensor with perfect fluid matter and electromagnetic field (12) reduces to (11c), when

$$\bar{\mu} = \mu \pm \phi^2, \quad p_{\parallel} = p \mp \phi^2, \quad p_{\perp} = p \pm \phi^2. \qquad (13a)$$

Therefore,

$$p_{\parallel} + 2p_{\perp} = 3p \pm \phi^2, \quad p_{\perp} - p_{\parallel} = \pm 2\phi^2. \qquad (13b)$$

Furthermore, it follows from (11a) and (13b) that (12) will also reduce to (11a) without energy flux ($q^a = 0$), when

$$3\bar{p} = 3p \pm \phi^2, \quad \Pi_{ab} = \pm 2\phi^2\left[\tfrac{1}{3}P_{ab} - n_a n_b\right]. \qquad (13c)$$

Finally, using (10), (11b), and (13c), for viscous fluids (with no energy flux) the following will hold if $\xi^a = \xi u^a$:

$$\eta = \pm[\phi^2/\sigma]. \qquad (14)$$

Note 2: The above equivalence is subject to the Rainchi–Misner–Wheeler[23] conditions. See also Refs. 12 and 13.

Conclusion 3: We have established that our present study is applicable to a large variety of physical problems in relativistic fluids.

For an application to cosmology (θ is not identically zero) we first mention that although shear-free solutions would retain the feature of isotropy of local motions, nonvanishing shear solutions are important as the red shift and the microwave background need no longer be isotropic. To illustrate this property, let a connecting vector X be drawn from points on one galactic world line to a neighboring world line. Then the vector $X^a = P^a{}_b X^b$, which lies in the instantaneous rest space of the first observer, has magnitude δl and direction n^a, where $n^a n_a = 1$ and $n^a u_a = 0$. The rate of change of relative distance is[22]

$$(\dot{\delta l})/\delta l = \tfrac{1}{3}\theta + \sigma_{ab}n^a n^b. \qquad (15a)$$

The relative red shift of galaxies in the direction of n^a is

$$d\lambda/\lambda = (\tfrac{1}{3}\theta + \sigma_{ab}n^a n^b + \dot{u}_a n^a)\delta l. \qquad (15b)$$

The rate of change of direction is

$$P^a{}_b(n^b)^{\cdot} = [\omega^a{}_b + \sigma^a{}_b - (\sigma_{cd}n^c n^d)P^a{}_b]n^b. \qquad (15c)$$

It is known[24] that there is a preferred direction [except for some degenerate cases, such as Friedmann–Robertson–Walker (FRW) models] so that neighboring clusters of galaxies appear to be instantaneously fixed in a local inertial frame. For shear-free ($\sigma_{ab} = 0$) and irrotational ($\omega_{ab} = 0$) fluids, this is clearly true of all directions. However, for shear-free and rotational fluids ($\omega_{ab} \neq 0$), this property holds only in the direction of the vorticity vector ω^a. Observe that the above general results are compatible with Theorem 1. However, for a possible physical interpretation with respect to the (Conf C) mappings, we first state the following proposition.

Proposition: If a fluid space-time admits timelike (Conf C) with symmetry vector $\xi^a = \xi u^a$ ($\xi > 0$), then

$$(\dot{\delta l})/\delta l = \tfrac{1}{3}(\theta + \sqrt{3}\sigma),$$
$$d\lambda/\lambda = \left[\tfrac{1}{3}(\theta + \sqrt{3}\sigma) + \dot{u}_a n^a\right]\delta l,$$
$$P^a{}_b(n^b)^{\cdot} = \omega^a{}_b n^b + (1/2\sqrt{3})\sigma n^a.$$

Conclusion 4: If $\sigma = 0$, then we recover the results on shear-free cosmologies (see Collins, Ref. 25).

For $\sigma \neq 0$, we observe that all the three observational aspects are anisotropic. For the irrotational ($\omega_{ab} = 0$) fluids, the motion of neighboring galaxies is in the direction of n^a, but the rotational ($\omega_{ab} \neq 0$) cosmological models are, in general, degenerate as there are no preferred directions. In particular, if n^a is chosen parallel to the vorticity vector ω^a, then the preferred direction coincides with n^a.

V. CONCLUDING REMARKS

It is evident from this paper that, whereas the timelike conformal motions inherit shear-free property, the conformal collineation mappings with timelike symmetry vector parallel to the velocity vector play an important role in the study of anisotropic fluids. The essential distinct property of this congruence is the *nonvanishing of the shear tensor.*

For comparative study, we have considered only one (Theorem 1) of the several other results on timelike congruences and (Conf M). For example, it is known[26] that two different timelike conformal motions cannot have the same streamlines. There are quite a few useful results on the propagation equations for tensor quantities along a curve of the congruence. *Thus we recommend further study on the physically relevant aspects of timelike congruences as related to conformal collineations.*

In another direction, the investigation of spacelike (null) conformal collineation should provide a natural counterpart to the study of spacelike (null) conformal motions, for which the notion of shear, expansion, vorticity, and acceleration may be defined in a manner similar to that outlined in Ref. 21. For a recent study on spacelike conformal motions, see Ref. 19.

Finally, it must be mentioned that the subject matter presented in this paper is a preliminary and tentative attempt that undoubtedly leaves much room for improvement. However, it does illustrate progress in providing some groundwork for further study on space-time with nonvanishing shear. It is hoped that the mathematical example of an affine conformal vector (constructed in Sec. III) may be useful in gaining insight into the really difficult question of finding a physical example of a conformal collineation other than conformal motion.

ACKNOWLEDGMENTS

I wish to express my gratitude to D. P. Mason, R. Maartens, and L. Herrera who have shown interest in this work through comments and correspondence.

My thanks, also, go to the Natural Sciences and Engineering Research Council of Canada for a research grant.

[1]H. Weyl, Gottingen Nachr. **1921**, 99.
[2]R. Penrose, Proc. R. Soc. London A Ser. **284**, 159 (1965).
[3]C. N. Yang, Phys. Rev. D **16**, 330 (1977).
[4]L. Herrera, J. Jimenez, L. Ponce de Leon, M. Esculpi, and V. Galino, J. Math. Phys. **25**, 3274 (1984).
[5]L. Herrera and J. Ponce de Leon, J. Math. Phys. **26**, 778, 2018, 2847 (1985).
[6]M. Esculpi and L. Herrera, J. Math. Phys. **27**, 2087 (1986).
[7]J. Carot and L. Mas, J. Math. Phys. **27**, 2336 (1986).
[8]M. Cahil and A. H. Taub, Commun. Math. Phys. **21**, 1 (1971).
[9]D. M. Eardley, Commun. Math. Phys. **37**, 287 (1974).
[10]J. R. Wainwright, in *Axisymmetric Systems*, edited by M. A. H. MacCallum (Cambridge U.P., Cambridge, 1985).
[11]K. L. Duggal and R. Sharma, J. Math. Phys. **27**, 2511 (1986).
[12]B. O. J. Tupper, J. Math. Phys. **22**, 2666 (1981).
[13]S. R. Maiti and S. N. Das, J. Math. Phys. **26**, 2840 (1985).
[14]J. L. Synge, University of Toronto Studies, Series No. 1, 1953.
[15]Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds and Physics* (North-Holland, Amsterdam, 1977).
[16]B. O'Neil, *Semi-Riemannian Geometry with Applications to Relativity* (Academic, New York, 1983).
[17]S. W. Hawking and G. F. S. Ellis, *The Large Scale Structure of Space-time* (Cambridge U.P., Cambridge, 1973).
[18]A. K. Raychaudhuri, Phys. Rev. **106**, 172 (1955).
[19]D. P. Mason and M. Tsamparlis, J. Math. Phys. **26**, 2881 (1985).
[20]D. R. Oliver and W. R. Davis, Gen. Relativ. Gravit. **8**, 905 (1977).
[21]For a timelike congruence of curves with unit tangent vector u^a, the vorticity tensor ω_{ab}, the expansion θ, the acceleration vector $\dot{u}^a$, and the shear tensor σ_{ab} are defined by

$$\omega_{ab} = P^c{}_a P^d{}_b u_{[c;d]},$$

$$\theta = P^{ab} u_{a;b}, \quad \dot{u}^a = u^a{}_{;b} u^b,$$

$$\sigma_{ab} = P^c{}_a P^d{}_b u_{(c;d)} - (\theta/3) P_{ab}.$$

[22]G. F. R. Ellis, "Relativistic cosmology," in *General Relativity and Gravitation*, XLIII Enrico Fermi Summer School Proceedings, edited by R. K. Sachs (Academic, New York, 1971).
[23]C. W. Misner and J. A. Wheeler, Ann. Phys. (NY) **2**, 525 (1957).
[24]J. M. Stewart, Mon. Not. R. Astron. Soc. **145**, 347 (1969).
[25]B. Collins, Can. J. Phys. **9**, 191 (1986).
[26]G. Williams, J. Math. Anal. Appl. **22**, 646 (1968).

Rotating rigid motion in general relativity

D. P. Mason
Centre for Nonlinear Studies and Department of Computational and Applied Mathematics, University of the Witwatersrand, Johannesburg 2050, South Africa
C. A. Pooe
Department of Mathematics, University of Bophuthatswana, Mafikeng 8670, Bophuthatswana

(Received 25 February 1987; accepted for publication 15 July 1987)

Kinematic and dynamic expressions are derived for the Lie derivative of vorticity along a particle world line in a rigid motion. It is found that the evolution of vorticity in a rigid motion is governed by the electric part of the Weyl tensor. Necessary and sufficient kinematic and dynamic conditions are established for a rotating rigid motion to be isometric.

I. INTRODUCTION

In this paper we study the kinematic and dynamic properties of rotating rigid motion in general relativity.

A definition of rigid motion in Minkowski space-time was proposed by Born[1] for rectilinear rigid motion and then independently by Herglotz[2] and Noether[3] for general rigid motion. Salzman and Taub[4] suggested that the same definition could be used for rigid motion in a curved space-time: A body is called rigid if the distance between every neighboring pair of particles measured orthogonal to the world line of either of them, remains constant along the world line.

In a rigid motion of a continuous medium, the shear tensor σ_{ab} and the rate of expansion θ both vanish and therefore the only nonzero kinematic quantities besides the kinematic four-velocity u^a are the vorticity vector and vorticity tensor, ω^a and ω_{ab}, and the four-acceleration vector $\dot{u}^a$.

Rayner,[5] Pirani and Williams,[6] and Trautman[7] have shown with the aid of Einstein's field equations that the angular velocity (vorticity) of a rigid heavy body, which contributes to the gravitational field, and of a rigid test body *in vacuo*, are constant in magnitude along any particle world line of the body. The condition for a heavy body is that the kinematic four-velocity u^a, which occurs in the equations of rigid motion, is the same as the dynamic four-velocity u^a_D, which is the timelike eigenvector of the energy-momentum tensor.[6–8] (The kinematic and dynamic four-velocities are characterized, respectively, by the properties that an observer with four-velocity u^a sees no particle flux density while an observer with four-velocity u^a_D measures vanishing energy flux.) For a continuous medium such as a fluid, u^a is in general not equal to u^a_D and therefore q^a, the energy flux relative to u^a, does not in general vanish. By expressing the Riemann curvature tensor in terms of the Weyl and Ricci tensors we derive a new kinematic expression, and a new dynamic expression which does not assume $q^a = 0$, for the rate of change of the magnitude of the vorticity along a particle world line in a rigid motion. We also derive new kinematic and dynamic expressions for the Lie derivative of ω_a and ω_{ab} along a particle world line in a rigid motion. It is found that the evolution of vorticity in a rigid motion is governed by the "electric" part of the Weyl tensor E_{ab} and is independent of the "magnetic" part of the Weyl tensor H_{ab}.

A motion is isometric if there exists a Killing vector field everywhere tangent to the world lines of the material particles. Every isometric motion is rigid but a rigid motion is isometric if and only if the acceleration four-vector $\dot{u}_a$ is the gradient of a scalar. Herglotz[2] and Noether[3] have shown that every rotating rigid motion in flat space-time is isometric and this result is sometimes referred to as the Herglotz–Noether theorem. Williams[9] has extended the Herglotz–Noether theorem to space-times of constant curvature. William's results, like that of Herglotz and Noether, is purely kinematical. Pirani and Williams[6] and Trautman,[7] however, have shown that there exist curved space-times which admit nonisometric rotating rigid motions. Boyer[10] has derived sufficient, but not necessary, conditions for the Herglotz–Noether theorem to be valid in curved space-times. We will derive here necessary and sufficient kinematic conditions for a rotating rigid motion to be isometric. We then apply Einstein's field equations to investigate the dynamic restrictions imposed on these kinematic conditions.

In Sec. II the equations of rigid motion are briefly reviewed. Section III is concerned with the vorticity of a rigid motion. Kinematic and dynamic expressions are derived for the Lie derivative of the magnitude of the vorticity ω and for ω_a and ω_{ab} along a particle world line. In Sec. IV, necessary and sufficient kinematic and dynamic conditions for a rotating rigid motion to be isometric are established. Concluding remarks are made in Sec. V.

The notation of Ellis[11,12] will be followed throughout.[13]

II. EQUATIONS OF RIGID MOTION

In this section we review the equations of rigid motion. This serves also to introduce notation and define quantities required later.

We first establish necessary and sufficient kinematic conditions for a motion to be rigid.[6,7] The equations of the world lines of the particles of a body or continuous medium may be written as $x^a = x^a(y^\alpha,\tau)$, where $\{y^\alpha\}$ are the coordinates of the particles in some arbitrarily chosen space section S, and τ denotes proper time measured along the world lines from the initial surface S. The kinematic four-velocity u^a is defined by

$$u^a = \left.\frac{\partial x^a}{\partial \tau}\right|_{y^\alpha = \text{const}} \tag{2.1}$$

 0022-2488/87/112705-09$02.50

and satisfies $u_a u^a = -1$. Consider two neighboring world lines $C: x^a = x^a(y^\alpha, \tau)$ and $C': x^a = x^a(y^\alpha + \delta y^\alpha, \tau)$. The connecting vector δx^a joins the points (y^α, τ) and $(y^\alpha + \delta y^\alpha, \tau)$ on the world lines C and C' for all proper time τ. It can be shown that[11,12]

$$\mathscr{L}_u \delta x^a = 0, \tag{2.2}$$

where $\mathscr{L}_u$ stands for the Lie derivative along u^a. The relative position vector is

$$\delta x_\perp^a = h^a_{\ b} \delta x^b, \tag{2.3}$$

where

$$h_{ab} = g_{ab} + u_a u_b \tag{2.4}$$

is the projection tensor onto the three-plane orthogonal to u^a. The orthogonal distance from C to C' is

$$\delta l_\perp^2 = g_{ab} \delta x_\perp^a \delta x_\perp^b = h_{ab} \delta x^a \delta x^b. \tag{2.5}$$

For a rigid motion, $\delta l_\perp$ remains constant along C; thus

$$(h_{ab} \delta x^a \delta x^b)^{\cdot} = 0, \tag{2.6}$$

which may be rewritten with the aid of (2.2) as

$$(\mathscr{L}_u h_{ab}) \delta x^a \delta x^b = 0. \tag{2.7}$$

Also, since $h_{ab} u^b = 0$ and $\mathscr{L}_u u^b \equiv 0$, we have

$$(\mathscr{L}_u h_{ab}) u^b = 0. \tag{2.8}$$

Since (2.7) must hold for all world lines C' near to C it follows from (2.7) and (2.8) that a motion is rigid if and only if

$$\mathscr{L}_u h_{ab} = 0. \tag{2.9}$$

Since[11,12]

$$u_{a;b} = \sigma_{ab} + (\theta/3) h_{ab} + \omega_{ab} - \dot{u}_a u_b, \tag{2.10}$$

where σ_{ab} is the shear tensor, θ is the rate of expansion, ω_{ab} is the vorticity tensor, and $\dot{u}_a$ is the acceleration vector, it can easily be verified that

$$\mathscr{L}_u h_{ab} = 2\sigma_{ab} + \tfrac{2}{3}\theta h_{ab}, \tag{2.11}$$

and therefore (2.9) is satisfied and the motion is rigid if and only if

$$\sigma_{ab} = 0 \quad \text{and} \quad \theta = 0. \tag{2.12}$$

The nonzero kinematic quantities in a rigid motion are therefore u^a, ω_{ab}, and $\dot{u}^a$.

The integrability conditions for the equations of rigid motion (2.9) have been derived by Pirani and Williams.[6] The first integrability condition is

$$\mathscr{L}_u(-3\omega_{ab}\omega_{cd} + \perp R_{abcd}) = 0, \tag{2.13}$$

where R_{abcd} is the Riemann curvature tensor of space-time. The projection symbol $\perp$ before a tensor expression denotes that after any indicated contractions in that expression have been performed each remaining free index is projected with $h_a^{\ b}$ (Ref. 13). Equation (2.13) can conveniently be derived by projecting the identity[14]

$$\mathscr{L}_u R^a_{\ bcd} = 2(\mathscr{L}_u \Gamma^a_{b[d})_{;c]} \tag{2.14}$$

with $h^a_{\ b}$ on all indices, where Γ^a_{bc} is the Christoffel symbol of the second kind. An outline of the derivation of (2.13) is given in Appendix A.

Consider now any covariant tensor $T_{ab\cdots c}$ which is orthogonal to u^a on all of its indices and which satisfies

$$\mathscr{L}_u T_{ab\cdots c} = 0. \tag{2.15}$$

Then if the motion is rigid[6,7]

$$\mathscr{L}_u(\perp T_{ab\cdots c;d}) = 0. \tag{2.16}$$

An outline of the derivation of this result is given in Appendix B for the case of a covariant vector T_a. The proof extends immediately to higher-order covariant tensors. It therefore follows from (2.13) and (2.16) that a second integrability condition of the equations of rigid motion is

$$\mathscr{L}_u(\perp(-3\omega_{ab}\omega_{cd} + \perp R_{abcd})_{;e}) = 0, \tag{2.17}$$

and by performing higher-order derivatives further integrability conditions can be written down.

The physical significance of the integrability condition (2.13) can be seen by considering the quotient space (space-time)/(world lines), i.e., the space of dimension 3 consisting of space-time with identification of points lying on the same member of the congruence of curves generated by u^a (Refs. 6, 7, and 15). The tensors $T_{ab\cdots c}$ in this quotient space are orthogonal to u^a on all indices and satisfy (2.15). When the motion is rigid the quotient space has a well defined metric tensor h_{ab}. We define the covariant derivative ${}^3\nabla_d$ in the quotient space by

$${}^3\nabla_d T_{ab\cdots c} = h^r_{\ a} h^s_{\ b} \cdots h^t_{\ c} h^f_{\ d} T_{rs\cdots t;f} = \perp T_{ab\cdots c;d}. \tag{2.18}$$

The three-space Riemann tensor ${}^3R_{abcd}$ of the quotient space is defined by the three-space Ricci identities

$${}^3\nabla_c {}^3\nabla_b T_a - {}^3\nabla_b {}^3\nabla_c T_a = {}^3R_{rabc} T^r, \tag{2.19}$$

which hold for any vector T_a in the quotient space, i.e., for any vector T_a which satisfies

$$T_a u^a = 0, \quad \mathscr{L}_u T_a = 0. \tag{2.20}$$

In order to obtain ${}^3R_{abcd}$ we evaluate the left-hand side of (2.19). For a rigid motion

$${}^3\nabla_c {}^3\nabla_b T_a = \omega_{ac} h^s_{\ b} u^r T_{r,s} + \omega_{bc} h^r_{\ a} \dot{T}_r + \perp T_{a;bc}. \tag{2.21}$$

But, for a rigid motion, since $u^r T_r = 0$,

$$h^s_{\ b} u^r T_{r,s} = \omega_{br} T^r, \tag{2.22}$$

and since $\mathscr{L}_u T_r = 0$,

$$h^r_{\ a} \dot{T}_r = \omega_{ar} T^r. \tag{2.23}$$

Thus

$${}^3\nabla_c {}^3\nabla_b T_a = (\omega_{ac}\omega_{br} + \omega_{bc}\omega_{ar}) T^r + \perp T_{a;bc}. \tag{2.24}$$

Hence, with the aid of the Ricci identity applied to T_a,

$$T_{a;[bc]} = \tfrac{1}{2} R_{rabc} T^r, \tag{2.25}$$

it follows that

$${}^3\nabla_c {}^3\nabla_b T_a - {}^3\nabla_b {}^3\nabla_c T_a = (2\omega_{bc}\omega_{ar} + \omega_{ac}\omega_{br} - \omega_{ab}\omega_{cr} + \perp R_{rabc}) T^r. \tag{2.26}$$

But since $\omega_{ab} = -\omega_{ba}$ and $\omega_{ab} u^b = 0$, it can be verified that

$$\omega_{a[b}\omega_{cd]} = 0 \tag{2.27}$$

and therefore

$${}^3\nabla_c {}^3\nabla_b T_a - {}^3\nabla_b {}^3\nabla_c T_a = (-3\omega_{ra}\omega_{bc} + \perp R_{rabc}) T^r. \tag{2.28}$$

Since T_a is an arbitrary vector in the quotient space it follows from (2.19) and (2.28) that

$$^3R_{abcd} = -3\omega_{ab}\omega_{cd} + \perp R_{abcd}. \tag{2.29}$$

The integrability condition (2.13) may therefore be rewritten as

$$\mathscr{L}_u\, {}^3R_{abcd} = 0. \tag{2.30}$$

We define the three-space Ricci tensor $^3R_{ab}$ and the three-space Ricci scalar ^3R by

$$^3R_{ab} = h^{rs}\,{}^3R_{rasb}, \tag{2.31}$$

$$^3R = h^{ab}\,{}^3R_{ab}. \tag{2.32}$$

Since for a rigid motion

$$\mathscr{L}_u h^{ab} = 2\dot{u}^{(a}u^{b)} \tag{2.33}$$

and $^3R_{abcd}$ is orthogonal to u^a on all indices, we obtain from (2.30) the following two results, valid for a rigid motion:

$$\mathscr{L}_u\, {}^3R_{ab} = 0, \tag{2.34}$$

$$\mathscr{L}_u\, {}^3R = 0. \tag{2.35}$$

Finally we observe that (2.16) may be written as

$$\mathscr{L}_u ({}^3\nabla_d T_{ab\cdots c}) = 0, \tag{2.36}$$

which holds for a rigid motion and for tensors $T_{ab\cdots c}$ orthogonal to u^a on all indices and which satisfy (2.15).

III. VORTICITY IN A RIGID MOTION

This section is concerned with the evolution of vorticity in a rigid motion. Kinematic results will first be derived for the rate of change of the magnitude of the vorticity along a particle world line and also for the Lie derivative of the vorticity vector and the vorticity tensor along a particle world line. We will then impose Einstein's field equations to investigate the dynamic restrictions imposed by a rigid motion on vorticity.

A. Kinematic results

In order to apply (2.30), (2.34), and (2.35) we first derive kinematic expressions for $^3R_{abcd}$, $^3R_{ab}$, and ^3R. We have

$$^3R_{ab} = 3(\omega_a\omega_b - \omega^2 h_{ab}) + \perp R_{arbs}h^{rs}, \tag{3.1}$$

$$^3R = R_{abcd}h^{ac}h^{bd} - 6\omega^2, \tag{3.2}$$

where[16]

$$\omega^a = \tfrac{1}{2}\eta^{abcd}u_b\omega_{cd}, \tag{3.3}$$

$$\omega^2 = \omega_a\omega^a = \tfrac{1}{2}\omega_{ab}\omega^{ab}, \tag{3.4}$$

and in deriving (3.1) we used the identity

$$\omega_{at}\omega_b{}^t = \omega^2 h_{ab} - \omega_a\omega_b, \tag{3.5}$$

which is obtained using the inverse of (3.3),

$$\omega_{ab} = \eta_{abcd}\omega^c u^d. \tag{3.6}$$

We now express the Riemann curvature tensor R_{abcd} in terms of the Weyl tensor C_{abcd}, the Ricci tensor R_{ab}, and the Ricci scalar R; we have[11,12]

$$R_{ab}{}^{cd} = C_{ab}{}^{cd} + 2g_{[a}{}^{[c}R_{b]}{}^{d]} - \tfrac{1}{3}Rg_{[a}{}^{[c}g_{b]}{}^{d]}. \tag{3.7}$$

The Weyl tensor can be decomposed with respect to u^a as

$$C_{ab}{}^{cd} = -8u_{[a}E_{b]}{}^{[c}u^{d]} + 4g_{[a}{}^{c]}E_{b]}{}^{d]} + 2\eta_{abrs}u^rH^{s[c}u^{d]} + 2\eta^{cdrs}u_rH_{s[a}u_{b]}, \tag{3.8}$$

where the electric part E_{ab} and the magnetic part H_{ab} of the Weyl tensor are defined by

$$E_{ab} = C_{arbs}u^ru^s, \quad H_{ab} = \tfrac{1}{2}\eta_{arpq}C^{pq}{}_{bs}u^ru^s, \tag{3.9}$$

and satisfy

$$E_{ab} = E_{ba}, \quad E_{ab}u^b = 0, \quad E^a{}_a = 0;$$
$$H_{ab} = H_{ba}, \quad H_{ab}u^b = 0, \quad H^a{}_a = 0. \tag{3.10}$$

By substituting from (3.7) and (3.8) into (2.29), (3.1), and (3.2) a direct calculation gives

$$\begin{aligned} ^3R_{abcd} = & -3\omega_{ab}\omega_{cd} + 2h_{c[a}(E_{b]d} + \tfrac{1}{2}\perp R_{b]d}) \\ & - 2h_{d[a}(E_{b]c} + \tfrac{1}{2}R_{b]c}) \\ & - (R/6)(h_{ac}h_{bd} - h_{bc}h_{ad}), \end{aligned} \tag{3.11}$$

$$\begin{aligned} ^3R_{ab} = & 3\omega_a\omega_b + E_{ab} + \tfrac{1}{2}\perp R_{ab} \\ & + (\tfrac{1}{2}R_{st}u^su^t + \tfrac{1}{6}R - 3\omega^2)h_{ab}, \end{aligned} \tag{3.12}$$

$$^3R = 2R_{st}u^su^t + R - 6\omega^2. \tag{3.13}$$

The three-space curvature tensors $^3R_{abcd}$, $^3R_{ab}$, and ^3R are independent of the magnetic part of the Weyl tensor, H_{ab}.

A simplification can be achieved by rewriting (3.11)–(3.13) in terms of the Einstein tensor G_{ab}. Since $G_{ab} = R_{ab} - \tfrac{1}{2}Rg_{ab}$ we have

$$R^a{}_a = -G^a{}_a, \tag{3.14}$$

$$R_{st}u^su^t = G_{st}u^su^t + \tfrac{1}{2}G^a{}_a, \tag{3.15}$$

$$\perp R_{ab} = G_{ab} - \tfrac{1}{3}(h^{st}G_{st})h_{ab} + (\tfrac{1}{3}G_{st}u^su^t - \tfrac{1}{6}G^a{}_a)h_{ab}, \tag{3.16}$$

and (3.11)–(3.13) become

$$\begin{aligned} ^3R_{abcd} = & -3\omega_{ab}\omega_{cd} \\ & + 2h_{c[a}(E_{b]d} + \tfrac{1}{2}(\perp G_{b]d} - \tfrac{1}{3}(h^{st}G_{st})h_{b]d})) \\ & - 2h_{d[a}(E_{b]c} + \tfrac{1}{2}(\perp G_{b]c} - \tfrac{1}{3}(h^{st}G_{st})h_{b]c})) \\ & + \tfrac{1}{3}(G_{st}u^su^t)(h_{ac}h_{bd} - h_{bc}h_{ad}), \end{aligned} \tag{3.17}$$

$$\begin{aligned} ^3R_{ab} = & 3\omega_a\omega_b + E_{ab} + \tfrac{1}{2}(\perp G_{ab} - \tfrac{1}{3}(h^{st}G_{st})h_{ab}) \\ & + (\tfrac{2}{3}G_{st}u^su^t - 3\omega^2)h_{ab}, \end{aligned} \tag{3.18}$$

$$^3R = 2G_{st}u^su^t - 6\omega^2. \tag{3.19}$$

Like E_{ab}, the tensor $\perp G_{ab} - \tfrac{1}{3}(h^{st}G_{st})h_{ab}$ is trace-free, symmetric, and orthogonal to u^a (Ref. 17). In place of the two scalar quantities $R_{st}u^su^t$ and R in (3.11)–(3.13), Eqs. (3.17)–(3.19) depend on the one-scalar quantity $G_{st}u^su^t$. The scalar $G_{st}u^su^t$ and the tensor $\perp G_{ab} - \tfrac{1}{3}(h^{st}G_{st})h_{ab}$ are directly related to dynamic quantities through Einstein's field equations [see (3.46) and (3.47) below].

In order to simplify subsequent expressions we define

$$\mathscr{E}_{ab} = E_{ab} + \tfrac{1}{2}(\perp G_{ab} - \tfrac{1}{3}(h^{st}G_{st})h_{ab}), \tag{3.20}$$

which satisfies

$$\mathscr{E}_{ab} = \mathscr{E}_{ba}, \quad \mathscr{E}_{ab}u^b = 0, \quad \mathscr{E}^a{}_a = 0. \tag{3.21}$$

Theorem 3.1: If the motion is rigid

(i) $$\dot{\omega} = -(1/4\omega^3)(\omega^c\omega^d \mathscr{L}_u \mathscr{E}_{cd}); \qquad (3.22)$$

$$(G_{st}u^s u^t)^{\cdot} = -(3/2\omega^2)(\omega^c\omega^d \mathscr{L}_u \mathscr{E}_{cd}); \qquad (3.23)$$

(ii) $$\mathscr{L}_u \omega_a = (1/12\omega^4)(\omega^c\omega^d \mathscr{L}_u \mathscr{E}_{cd})\omega_a - (1/3\omega^2)\omega^b \mathscr{L}_u \mathscr{E}_{ab}; \qquad (3.24)$$

(iii) $$\mathscr{L}_u \omega_{ab} = (1/12\omega^4)(\omega^c\omega^d \mathscr{L}_u \mathscr{E}_{cd})\omega_{ab} + (2/3\omega^2)\omega_{[a}{}^t \mathscr{L}_u \mathscr{E}_{b])t}; \qquad (3.25)$$

(iv) $$\dot{\omega} = 0 \Leftrightarrow \omega^c\omega^d \mathscr{L}_u \mathscr{E}_{cd} = 0, \qquad (3.26)$$

$$\mathscr{L}_u \omega_a = 0 \Leftrightarrow \mathscr{L}_u \mathscr{E}_{ab} = 0 \Leftrightarrow \mathscr{L}_u \omega_{ab} = 0. \qquad (3.27)$$

Proof: (i) By substituting from (3.19) into (2.35) we obtain

$$(G_{st}u^s u^t)^{\cdot} = 6\omega\dot{\omega}. \qquad (3.28)$$

Also, it follows from (2.34) and (3.18) that

$$3\omega_a \mathscr{L}_u \omega_b + 3\omega_b \mathscr{L}_u \omega_a + \mathscr{L}_u \mathscr{E}_{ab} + \left[\tfrac{2}{3}(G_{st}u^s u^t)^{\cdot} - 6\omega\dot{\omega}\right]h_{ab} = 0, \qquad (3.29)$$

and by contracting (3.29) with ω^b we obtain

$$3\omega_a(\omega^b \mathscr{L}_u \omega_b) + 3\omega^2 \mathscr{L}_u \omega_a + \omega^b \mathscr{L}_u \mathscr{E}_{ab} + \left[\tfrac{2}{3}(G_{st}u^s u^t)^{\cdot} - 6\omega\dot{\omega}\right]\omega_a = 0. \qquad (3.30)$$

But by taking the Lie derivative along u^a of $\omega_a\omega^a = \omega^2$ and using for a rigid motion the identity

$$\mathscr{L}_u g^{ab} = 2\dot{u}^{[a}u^{b]} \qquad (3.31)$$

it can be shown for a rigid motion that

$$\omega^a \mathscr{L}_u \omega_a = \omega\dot{\omega}. \qquad (3.32)$$

Equation (3.30) therefore simplifies to

$$3\omega^2 \mathscr{L}_u \omega_a + \omega^b \mathscr{L}_u \mathscr{E}_{ab} + \left[\tfrac{2}{3}(G_{st}u^s u^t)^{\cdot} - 3\omega\dot{\omega}\right]\omega_a = 0, \qquad (3.33)$$

and by contracting (3.33) with ω^a and using (3.32) we find that

$$(G_{st}u^s u^t)^{\cdot} = -(3/2\omega^2)(\omega^c\omega^d \mathscr{L}_u \mathscr{E}_{cd}). \qquad (3.34)$$

Equation (3.22) follows immediately from (3.28) and (3.34).

(ii) Equation (3.24) follows directly from (3.33), (3.22), and (3.23).

(iii) Equation (3.25) can be derived in two ways. First, (3.25) can be obtained using (3.24). Since $\omega_{ab} = \eta_{abcd}\omega^c u^d$ and

$$\mathscr{L}_u \eta_{abcd} = \theta\eta_{abcd} = 0, \qquad (3.35)$$

$$\mathscr{L}_u u^d \equiv 0, \qquad (3.36)$$

it follows with the aid of (3.31) that for a rigid motion

$$\mathscr{L}_u \omega_{ab} = \eta_{ab}{}^{cd}(\mathscr{L}_u \omega_c)u_d. \qquad (3.37)$$

By substituting from (3.24) into (3.37) and using the identity

$$\eta^{abcd}\eta_{efgh} = -4!\delta_e^{[a}\delta_f^b\delta_g^c\delta_h^{d]} \qquad (3.38)$$

Eq. (3.25) may be derived.

Alternatively, (3.25) may be obtained from (2.30) and (3.17). If (3.17) is substituted into (2.30), we obtain

$$-3\omega_{ab}\mathscr{L}_u \omega_{cd} - 3\omega_{cd}\mathscr{L}_u \omega_{ab} + 2h_{c[a}\mathscr{L}_u \mathscr{E}_{b]d} - 2h_{d[a}\mathscr{L}_u \mathscr{E}_{b]c} + \tfrac{1}{3}(G_{st}u^s u^t)^{\cdot}(h_{ac}h_{bd} - h_{bc}h_{ad}) = 0. \qquad (3.39)$$

But, by taking the Lie derivative along u^a of $\omega_{cd}\omega^{cd} = 2\omega^2$ and using (3.31) it can be shown for a rigid motion that

$$\omega^{cd}\mathscr{L}_u \omega_{cd} = 2\omega\dot{\omega}. \qquad (3.40)$$

Hence, by contracting (3.39) with ω^{cd} and using (3.40) we find that

$$6\omega^2 \mathscr{L}_u \omega_{ab} = \left[\tfrac{2}{3}(G_{st}u^s u^t)^{\cdot} - 6\omega\dot{\omega}\right] + 4\omega_{[a}{}^d \mathscr{L}_u \mathscr{E}_{b]d}. \qquad (3.41)$$

Equation (3.25) follows immediately with the aid of (3.22) and (3.23).

(iv) The result (3.26) is a direct consequence of (3.22). To show that $\mathscr{L}_u \omega_a = 0 \Leftrightarrow \mathscr{L}_u \mathscr{E}_{ab} = 0$, we first observe that if $\mathscr{L}_u \mathscr{E}_{ab} = 0$ then $\mathscr{L}_u \omega_a = 0$ from (3.24). Conversely, if $\mathscr{L}_u \omega_a = 0$ then $\dot{\omega} = 0$ by (3.32) and $(G_{st}u^s u^t)^{\cdot} = 0$ by (3.28) and therefore $\mathscr{L}_u \mathscr{E}_{ab} = 0$ by (3.29). To establish the remaining condition in (3.27) we can either show that $\mathscr{L}_u \omega_{ab} = 0 \Leftrightarrow \mathscr{L}_u \omega_a = 0$ or $\mathscr{L}_u \omega_{ab} = 0 \Leftrightarrow \mathscr{L}_u \mathscr{E}_{ab} = 0$. The former results follows from (3.37) and the inverse of (3.37),

$$\mathscr{L}_u \omega_a = \tfrac{1}{2}\eta_a{}^{bcd}u_b \mathscr{L}_u \omega_{cd}, \qquad (3.42)$$

which can be derived by operating on (3.37) with $\eta^{rsab}u_s$ and using

$$\mathscr{L}_u h_a{}^b = \dot{u}_a u^b. \qquad (3.43)$$

To establish the latter result we note that if $\mathscr{L}_u \mathscr{E}_{ab} = 0$ then $\mathscr{L}_u \omega_{ab} = 0$ by (3.25), and conversely by contracting (3.39) with h^{ac} and using (2.33) and (3.43), it is easily verified with the aid of (3.21) that $\mathscr{L}_u \omega_{ab} = 0 \Rightarrow \mathscr{L}_u \mathscr{E}_{ab} = 0$. □

B. Dynamic results

We now impose Einstein's field equations,

$$G_{ab} + \Lambda g_{ab} = T_{ab}, \qquad (3.44)$$

where Λ is the cosmological constant. The energy-momentum tensor T_{ab} may be decomposed with respect to the kinematic four-velocity u^a as

$$T_{ab} = \mu u_a u_b + p h_{ab} + 2q_{(a}u_{b)} + \pi_{ab}, \qquad (3.45)$$

where μ is the total energy density measured by an observer with four-velocity u^a, q^a is the energy flux relative to u^a $(q_a u^a = 0)$, p is the isotropic pressure, and π_{ab} is the trace-free anisotropic stress tensor ($\pi_{ab} = \pi_{ba}$, $\pi_{ab}u^b = 0$, $\pi^a{}_a = 0$). We have from (3.44) and (3.45),

$$G_{st}u^s u^t = \mu + \Lambda, \qquad (3.46)$$

$$\perp G_{ab} - \tfrac{1}{3}(h^{st}G_{st})h_{ab} = \pi_{ab}. \qquad (3.47)$$

Equations (3.22)–(3.25) become

$$\dot{\omega} = -(1/4\omega^3)\omega^c\omega^d \mathscr{L}_u(E_{cd} + \pi_{cd}), \qquad (3.48)$$

$$\dot{\mu} = -(3/2\omega^2)\omega^c\omega^d \mathscr{L}_u(E_{cd} + \pi_{cd}), \qquad (3.49)$$

$$\mathscr{L}_u \omega_a = (1/12\omega^4)(\omega^c\omega^d \mathscr{L}_u(E_{cd} + \pi_{cd}))\omega_a - (1/3\omega^2)\omega^b \mathscr{L}_u(E_{ab} + \pi_{ab}), \qquad (3.50)$$

$$\mathscr{L}_u\omega_{ab} = (1/12\omega^4)(\omega^c\omega^d\mathscr{L}_u(E_{cd}+\pi_{cd}))\omega_{ab} + (2/3\omega^2)\omega_{[a}{}^t\mathscr{L}_u(E_{b]t}+\pi_{b]t}). \quad (3.51)$$

An alternative expression for $\dot{\mu}$ is given by the energy conservation equation along a particle world line[11,12]:

$$\dot{\mu} + (\mu+p)\theta + \pi_{ab}\sigma^{ab} + q^a{}_{;a} + q_a\dot{u}^a + 0, \quad (3.52)$$

which simplifies for a rigid motion to

$$\dot{\mu} = -(q^a{}_{;a} + q_a\dot{u}^a). \quad (3.53)$$

Hence

$$\omega^c\omega^d\mathscr{L}_u(E_{cd}+\pi_{cd}) = (2\omega^2/3)(q^a{}_{;a}+q_a\dot{u}^a), \quad (3.54)$$

and

$$\dot{\omega} = -(1/6\omega)(q^a{}_{;a}+q_a\dot{u}^a). \quad (3.55)$$

The foregoing results apply in general to any rigid motion satisfying Einstein's field equations. We now consider two special cases, the rotation of a rigid heavy body and the rigid rotation of a fluid. In the first case $q^a=0$ and in the second $\pi_{ab}=0$.

The condition for a heavy body is that the kinematic four-velocity u^a is the same as the dynamic four-velocity u^a_D which is the timelike eigenvector of T^{ab} (Refs. 6–8),

$$T^{ab}u_{Db} = -\mu_D u^a_D, \quad (3.56)$$

where μ_D is the energy density of the body as measured by an observer with four-velocity u^a_D and is the minimum of the energy densities of the body measured by all possible observers. When $u^a = u^a_D$, it follows from (3.45) and (3.56) that $q^a=0$. Thus

$$\dot{\omega}=0, \quad (3.57)$$

$$\dot{\mu}=0, \quad (3.58)$$

$$\mathscr{L}_u\omega_a = -(1/3\omega^2)\omega^b\mathscr{L}_u(E_{ab}+\pi_{ab}), \quad (3.59)$$

$$\mathscr{L}_u\omega_{ab} = (2/3\omega^2)\omega_{[a}{}^t\mathscr{L}_u(E_{b]t}+\pi_{b]t}). \quad (3.60)$$

For a rigid test body *in vacuo*, $T_{ab}=0$ and (3.57)–(3.60) apply with $\pi_{ab}=0$. Equation (3.57) was first derived by Rayner[5]: for a rigid heavy body or a rigid test body *in vacuo*, the angular velocity is constant in magnitude along any particle world line of the body. Equation (3.55) gives the generalization of this result for the rotating rigid motion of a system with $q^a\neq 0$.

Consider now the rigid rotation of a fluid. The phenomenological equation of state

$$\pi_{ab} = -\lambda\sigma_{ab}, \quad \lambda\geqslant 0, \quad (3.61)$$

where λ is the coefficient of shear viscosity, is necessary if the rate of production of entropy is never negative.[11,12] Since $\sigma_{ab}=0$ in a rigid flow, it follows that $\pi_{ab}=0$ if (3.61) is satisfied, and Eqs. (3.48)–(3.51) with $\pi_{ab}=0$ then apply. The evolution of μ and of the vorticity of the fluid are controlled entirely by the electric part of the Weyl tensor, which in turn is related to q^a through (3.54) with $\pi_{cd}=0$. In general for a fluid, $q^a\neq 0$, and therefore from (3.55), $\dot{\omega}\neq 0$.

IV. RIGID AND ISOMETRIC ROTATING MOTIONS

A motion is isometric if there exists a Killing vector field parallel to the kinematic four-velocity field u^a. Necessary and sufficient conditions for a motion to be isometric are[6,7]

$$\sigma_{ab}=0, \quad \theta=0, \quad \dot{u}_{[a;b]}=0. \quad (4.1)$$

We see from (2.12) and (4.1) that every isometric motion is rigid but a rigid motion is isometric if and only if $\dot{u}_a$ is the gradient of a scalar.

We now establish necessary and sufficient conditions for a rotating rigid motion to be isometric. As in Sec. III we first establish kinematic conditions. We then investigate the dynamic restrictions imposed by Einstein's field equations.

A. Kinematic results

Theorem 4.1: A rotating rigid motion is isometric if and only if

$$\mathscr{L}_uE_{ab}=0, \quad (4.2)$$

$$\mathscr{L}_u(\perp G_{ab} - \tfrac{1}{3}(h^{st}G_{st})h_{ab})=0, \quad (4.3)$$

$$\mathscr{L}_uH_{ab}=0. \quad (4.4)$$

Proof: We first note some results which are true in general, even for a motion which is not rigid. The propagation equation for the vorticity tensor is

$$h^c_a h^d_b(\dot{\omega}_{cd} - \dot{u}_{[c;d]}) - 2\sigma^d{}_{[a}\omega_{b]d} + \tfrac{2}{3}\theta\omega_{ab}=0, \quad (4.5)$$

which may be rewritten as

$$\mathscr{L}_u\omega_{ab} = h^c_a h^d_b\dot{u}_{[c;d]}. \quad (4.6)$$

Also,

$$\mathscr{L}_u\dot{u}_a = 2\dot{u}_{[a;b]}u^b. \quad (4.7)$$

It follows directly from (4.6) and (4.7) that

$$\dot{u}_{[a;b]} = \mathscr{L}_u\omega_{ab} + u_{[a}\mathscr{L}_u\dot{u}_{b]}. \quad (4.8)$$

Hence

$$\dot{u}_{[a;b]}=0 \Leftrightarrow \mathscr{L}_u\omega_{ab}=0 \quad\text{and}\quad \mathscr{L}_u\dot{u}_a=0, \quad (4.9)$$

which forms the basis of the following proof.

(i) Suppose first that the motion is rigid and that conditions (4.2)–(4.4) are satisfied. We prove that the motion is isometric by showing that $\dot{u}_{[a;b]}=0$.

It follows from (4.2) and (4.3) that $\mathscr{L}_uE_{ab}=0$ and therefore from (3.25),

$$\mathscr{L}_u\omega_{ab}=0. \quad (4.10)$$

In order to obtain an expression for $\mathscr{L}_u\dot{u}_a$, consider the constraint equation[11,12]

$$H_{ab} = 2\dot{u}_{(a}\omega_{b)} - h^s_a h^t_b(\omega_{(s}{}^{d;c} + \sigma_{(s}{}^{d;c})\eta_{t)fdc}u^f. \quad (4.11)$$

Equation (4.11) is purely kinematical. Einstein's field equations are not used in its derivation. For a rigid motion, $\sigma_{ab}=0$, and (4.11) may be rewritten in the form

$$H_{ab} = 2\omega_{(a}\dot{u}_{b)} - h_{p(a}(h^s_{b)}h^m_d h^n_c\omega_{sm;n})\eta^{pfdc}u_f. \quad (4.12)$$

We take the Lie derivative along u^a of (4.12). Since $\mathscr{L}_u\mathscr{E}_{ab}=0$ it follows from (3.24) that $\mathscr{L}_u\omega_a=0$. Also, since ω_{ab} is orthogonal to u^a on both indices and (4.10) is satisfied we have for a rigid motion by (2.16),

$$\mathscr{L}_u(h^s_b h^m_d h^n_c\omega_{sm;n}) = \mathscr{L}_u(\perp\omega_{bd;c})=0. \quad (4.13)$$

Further

$$\mathscr{L}_u\eta^{pfdc} = -\theta\eta^{pfdc}=0, \quad (4.14)$$

$$\mathscr{L}_u u_f = \dot{u}_f. \tag{4.15}$$

Thus

$$\mathscr{L}_u H_{ab} = 2\omega_{(a}\mathscr{L}_u \dot{u}_{b)} - h_{p(a}(h^t_{\ b)} h^m_{\ d} h^n_{\ c}\omega_{tm;n})\eta^{pfdc}\dot{u}_f. \tag{4.16}$$

But it can be verified by a direct calculation using (3.6) and (3.38) that for a rigid motion

$$h_{pa} h^t_{\ b} h^m_{\ d} h^n_{\ c}\omega_{tm;n}\eta^{pfdc}\dot{u}_f = 0, \tag{4.17}$$

and therefore

$$\mathscr{L}_u H_{ab} = 2\omega_{(a}\mathscr{L}_u \dot{u}_{b)}. \tag{4.18}$$

Equation (4.18) can be solved for $\mathscr{L}_u \dot{u}_b$ by contracting first with ω^b and then with ω^a:

$$\mathscr{L}_u \dot{u}_a = (1/\omega^2)(\omega^b \mathscr{L}_u H_{ab} - (1/2\omega^2)(\omega^c\omega^d \mathscr{L}_u H_{cd})\omega_a). \tag{4.19}$$

Hence by (4.4),

$$\mathscr{L}_u \dot{u}_a = 0, \tag{4.20}$$

and therefore from (4.9), (4.10), and (4.20), $\dot{u}_{[a;b]} = 0$. Thus the rigid motion is isometric.

(ii) Conversely, suppose that the motion is isometric. It is shown in Appendix C by a direct calculation that (4.2)–(4.4) are satisfied. We give here an alternative proof using the converse of results established in part (i).

Consider first condition (4.4). Since the motion is isometric, $\dot{u}_{[a;b]} = 0$, and therefore from (4.9),

$$\mathscr{L}_u \omega_{ab} = 0, \quad \mathscr{L}_u \dot{u}_a = 0. \tag{4.21}$$

Since $\mathscr{L}_u \omega_{ab} = 0$ it follows from (2.16) and (3.42) that

$$\mathscr{L}_u(\perp\omega_{ab;c}) = 0, \quad \mathscr{L}_u \omega_a = 0. \tag{4.22}$$

But we have seen that if a rigid motion satisfies conditions (4.22) then (4.18) holds and since also $\mathscr{L}_u \dot{u}_a = 0$, it follows directly that $\mathscr{L}_u H_{ab} = 0$.

Consider now the remaining two conditions, (4.2) and (4.3). Since $\mathscr{L}_u \omega_{ab} = 0$ we have by (3.27) that $\mathscr{L}_u \mathscr{E}_{ab} = 0$; thus

$$\mathscr{L}_u E_{ab} + \tfrac{1}{2}\mathscr{L}_u(\perp G_{ab} - \tfrac{1}{3}(h^{st}G_{st})h_{ab}) = 0. \tag{4.23}$$

In order to obtain a second equation relating the Lie derivatives of E_{ab} and $\perp G_{ab} - \frac{1}{3}(h^{st}G_{st})h_{ab}$ consider the shear propagation equation in kinematic form,

$$\begin{aligned} & h^c_{\ a} h^d_{\ b}(\dot{\sigma}_{cd} - \dot{u}_{(c;d)}) - \dot{u}_a\dot{u}_b + \omega_a\omega_b + \sigma_{at}\sigma^t_{\ b} \\ & + \tfrac{2}{3}\theta\sigma_{ab} + \tfrac{1}{3}(\dot{u}^c_{\ ;c} - \omega^2 - 2\sigma^2)h_{ab} \\ & + E_{ab} - \tfrac{1}{2}(\perp G_{ab} - \tfrac{1}{3}(h^{st}G_{st})h_{ab}) = 0. \end{aligned} \tag{4.24}$$

Equation (4.24) is derived by taking the symmetric trace-free part of the propagation equation for $\perp\dot{u}_{a;b}$ (Refs. 11 and 12) and without applying Einstein's field equations. For an isometric motion, (4.24) reduces to

$$\begin{aligned} & -\perp\dot{u}_{a;b} - \dot{u}_a\dot{u}_b + \omega_a\omega_b + \tfrac{1}{3}(\dot{u}^c_{\ ;c} - \omega^2)h_{ab} \\ & + E_{ab} - \tfrac{1}{2}(\perp G_{ab} - \tfrac{1}{3}(h^{st}G_{st})h_{ab}) = 0. \end{aligned} \tag{4.25}$$

We take the Lie derivative along u^a of (4.25). Since $\dot{u}_a u^a = 0$ and $\mathscr{L}_u \dot{u}_a = 0$, it follows from (2.16) that

$$\mathscr{L}_u(\perp\dot{u}_{a;b}) = 0 \tag{4.26}$$

and by contracting (4.26) with g^{ab} and using (3.31) we find that

$$\mathscr{L}_u(\dot{u}^c_{\ ;c}) = 0. \tag{4.27}$$

Also, since $\mathscr{L}_u \omega_{ab} = 0$, we have from (3.40) that $\dot{\omega} = 0$. The Lie derivative along u^a of (4.25) therefore yields

$$\mathscr{L}_u E_{ab} - \tfrac{1}{2}\mathscr{L}_u(\perp G_{ab} - \tfrac{1}{3}(h^{st}G_{st})h_{ab}) = 0. \tag{4.28}$$

Conditions (4.2) and (4.3) follow immediately from (4.23) and (4.28). □

The kinematic results of Herglotz[2] and Noether[3] and of Williams[9] that every rotating rigid motion in a flat space-time and in a space-time of constant curvature is isometric are special cases of Theorem 4.1. We now check that conditions (4.2)–(4.4) are satisfied in both cases.

(i) Flat space-time. Since $R_{abcd} = 0$ we have

$$E_{ab} = 0, \quad H_{ab} = 0, \quad \perp G_{ab} - \tfrac{1}{3}(h^{st}G_{st})h_{ab} = 0, \tag{4.29}$$

and therefore (4.2)–(4.4) are identically satisfied.

(ii) Space-time of constant curvature. For a space-time of constant curvature

$$R_{abcd} = K(g_{ac}g_{bd} - g_{ad}g_{bc}), \tag{4.30}$$

where K is a constant. Thus

$$R_{ab} = 3Kg_{ab}, \quad R = 12K, \quad G_{ab} = -3Kg_{ab}. \tag{4.31}$$

It is easily verified from (3.7) that $C_{abcd} = 0$. We find that (4.29) again holds and therefore that conditions (4.2)–(4.4) are identically satisfied. The result remains valid if K depends on x^a.

B. Dynamic results

We now impose Einstein's field equations. The following theorem follows immediately from (3.47) and Theorem 4.1.

Theorem 4.2: If Einstein's field equations (3.44) are satisfied, then a rotating rigid motion is isometric if and only if

$$\mathscr{L}_u E_{ab} = 0, \quad \mathscr{L}_u \pi_{ab} = 0, \quad \mathscr{L}_u H_{ab} = 0. \tag{4.32}$$

If, further, the phenomenological equation of state,

$$\pi_{ab} = -\lambda\sigma_{ab} \quad (\lambda \geqslant 0) \tag{4.33}$$

is satisfied then a rotating rigid motion is isometric if and only if

$$\mathscr{L}_u E_{ab} = 0, \quad \mathscr{L}_u H_{ab} = 0. \tag{4.34}$$

□

We have observed that the phenomenological equation (4.33) is necessary in a fluid if the rate of generation of entropy is never negative. In that case only two conditions, given by (4.34), are necessary and sufficient for a rotating rigid motion to be isometric.

If Einstein's field equations (3.44) are satisfied and the motion is rigid, then E_{ab} and H_{ab} satisfy the propagation equations[11]

$$\begin{aligned} & \mathscr{L}_u E_{ab} + E_{t(a}\omega_{b)}^{\ \ t} + h^c_{\ (a}\eta_{b)rst}u^r H_c^{\ s;t} - 2H^t_{\ (a}\eta_{b)rst}u^r\dot{u}^s \\ & = -\dot{u}_{(a}q_{b)} - \tfrac{1}{2}h^s_{\ a}h^t_{\ b}q_{(s;t)} \ \tfrac{1}{6}(q^t_{\ ;t} + \dot{u}_t q^t)h_{ab} \\ & \quad - \tfrac{1}{2}h^s_{\ a}h^t_{\ b}\dot{\pi}_{st} + \tfrac{1}{2}\pi_{t(a}\omega_{b)}^{\ \ t}, \end{aligned} \tag{4.35}$$

$$\begin{aligned} & \mathscr{L}_u H_{ab} + H_{t(a}\omega_{b)}^{\ \ t} - h^c_{\ (a}\eta_{b)rst}u^r E_c^{\ s;t} + 2E^t_{\ (a}\eta_{b)rst}u^r\dot{u}^s \\ & = \tfrac{1}{2}((\omega_t q^t)h_{ab} - 3\omega_{(a}q_{b)}) - \tfrac{1}{2}h^c_{\ (a}\eta_{b)rst}u^r\pi_c^{\ s;t}, \end{aligned} \tag{4.36}$$

and the constraint equations

$$h^a_r E^{rs}{}_{;t} h^t_s + 3H^a{}_b\omega^b = \tfrac{1}{3}\mu_{,b}h^{ab} - 3\omega^{ab}q_b - \tfrac{1}{2}h^a_c\pi^{cb}{}_{;b} + \tfrac{1}{2}\pi^{ab}\dot{u}_b, \tag{4.37}$$

$$h^a_r H^{rs}{}_{;t} h^t_s - 3E^a{}_b\omega^b = (\mu + p)\omega^a + \tfrac{1}{2}\eta^{arst}u_r q_{s;t} - \tfrac{1}{2}\pi^{ab}\omega_b. \tag{4.38}$$

Hence in general, $\mathscr{L}_u E_{ab}$ and $\mathscr{L}_u H_{ab}$ will be nonzero in a rigid motion.

The results of this section apply to *rotating* rigid motions only. Equations (4.35)–(4.38) can place restrictions on the vorticity of a rigid motion. Consider for instance a conformally flat space-time with $q^a = 0$ and $\pi^{ab} = 0$. The latter condition would be satisfied in a rigid flow of a fluid obeying the phenomenological equation (4.33). It follows from (4.38) that if $\mu + p \neq 0$ then $\omega_a = 0$. Hence if $\mu + p \neq 0$, Theorem 4.2 will not apply, although conditions (4.32) are identically satisfied, because the motion is irrotational. [We observe, however, that if further $p = p(\mu)$, then from (4.37)

$$p_{,b}h^{ba} = \frac{dp}{d\mu}\mu_{,b}h^{ba} = 0, \tag{4.39}$$

and therefore it follows from the momentum conservation equation[11,12]

$$(\mu + p)\dot{u}^a = -p_{,b}h^{ba} \tag{4.40}$$

that $\dot{u}^a = 0$ if $\mu + p \neq 0$. Therefore the rigid motion is isometric.] A space-time of constant curvature is conformally flat and using (3.44), (3.45), and (4.31) it can be verified that

$$\mu = 3K - \Lambda, \quad p = \Lambda - 3K, \quad q^a = 0, \quad \pi^{ab} \doteq 0. \tag{4.41}$$

However, since $\mu + p = 0$, the vorticity need not vanish and the theorem of Williams,[9] that every *rotating* rigid motion in a space-time of constant curvature is isometric, has application. For a space-time of constant curvature, (4.35)–(4.38) are identically satisfied, but it follows from (3.22), (3.24), and (3.25) that

$$\dot{\omega} = 0, \quad \mathscr{L}_u\omega_a = 0, \quad \mathscr{L}_u\omega_{ab} = 0. \tag{4.42}$$

Finally we observe that if a rigid motion is irrotational then $q^a = 0$. This follows immediately from the $(0,\alpha)$ field equations,[11,12]

$$q^a = h^{ab}(\tfrac{2}{3}\theta_{;b} - \sigma_{bc;d}h^{cd}) - \eta^{abcd}u_b(\omega_{c;d} + 2\omega_c\dot{u}_d). \tag{4.43}$$

V. CONCLUDING REMARKS

We did not use the second integrability condition (2.17). Instead we found it more convenient to apply the integrability condition (2.16) of Eq. (2.15) directly to the tensors ω_{ab} and $\dot{u}_a$ which are orthogonal to u^a on all indices and which had been shown to have vanishing Lie derivatives along u^a.

The Born criterion for rigid motion is much more restrictive than may have been expected from Newtonian theory. For instance, if $q^a = 0$ or for the class of rigid motions which are isometric, the vorticity of a rigid body is constant in magnitude along any particle world line of the body. A view sometimes taken, therefore, is that a less restrictive definition should be found. However, as pointed out by Pirani and Williams,[6] restrictions on the motion of a rigid body may be explained by the argument that the application of forces to the body in order to change its momentum or angular momentum must distort the body so that a body which is required to be rigid cannot be subject to such forces. Hence severe restrictions on $\dot{u}^a$ and ω^a are to be expected.

ACKNOWLEDGMENT

DPM would like to thank the CSIR Foundation for Research Development for an Operating Grant.

APPENDIX A: INTEGRABILITY CONDITION FOR RIGID MOTION

An outline is given of the derivation of the integrability condition for the equations of rigid motion,

$$\mathscr{L}_u(-3\omega_{ab}\omega_{cd} + \perp R_{abdc}) = 0, \tag{A1}$$

by starting from the identity

$$\mathscr{L}_u R^r{}_{sfg} = 2(\mathscr{L}_u\Gamma^r_{s[g})_{;f]}. \tag{A2}$$

Consider a rigid motion. Project on the indices of (A2) with $h_{ar}h^s_b h^f_c h^g_d$; this gives

$$h_{ar}h^s_b h^f_c h^g_d \mathscr{L}_u R^r{}_{sfg} = 2h_{ar}h^s_b h^f_c h^g_d(\mathscr{L}_u\Gamma^r_{s[g})_{;f]}. \tag{A3}$$

We first evaluate the left-hand side of (A3). Since

$$\mathscr{L}_u h^b_a = \dot{u}_a u^b \tag{A4}$$

and, for a rigid motion, $\mathscr{L}_u h_{ab} = 0$, we have

$$h_{ar}h^s_b h^f_c h^g_d \mathscr{L}_u R^r{}_{sfg} = \mathscr{L}_u(\perp R_{abcd}) + \dot{u}_b \perp R_{tacd}u^t + \dot{u}_c \perp R_{tdba}u^t + \dot{u}_d \perp R_{tcab}u^t. \tag{A5}$$

Consider next the right-hand side of (A3). Using the identity[14]

$$\mathscr{L}_u\Gamma^a_{bc} = \tfrac{1}{2}g^{at}[(\mathscr{L}_u g_{bt})_{;c} + (\mathscr{L}_u g_{ct})_{;b} - (\mathscr{L}_u g_{bc})_{;t}], \tag{A6}$$

where for a rigid motion

$$\mathscr{L}_u g_{ab} = -2\dot{u}_{(a}u_{b)} \tag{A7}$$

a direct calculation gives

$$\begin{aligned} h_{ar}h^s_b h^f_c h^g_d(\mathscr{L}_u\Gamma^r_{sg})_{;f} &= \omega_{ac}(-\perp\dot{u}_{(b;d)} - \dot{u}_b\dot{u}_d) + \omega_{bc}(-\dot{u}_{[a;d]} + \dot{u}_a\dot{u}_d) \\ &\quad + \omega_{dc}(-\perp\dot{u}_{[a;b]} + \dot{u}_a\dot{u}_b) - 2\omega_{a(b}\perp\dot{u}_{d);c} \\ &\quad + 2\dot{u}_{(b}\perp\omega_{d)a;c}. \end{aligned} \tag{A8}$$

In order to derive an expression for $\dot{u}_{a;b}$ we expand $\dot{u}_{a;b} = (u_{a;t}u^t)_{;b}$ and use the Ricci identity applied to u_a,

$$u_{a;[bc]} = \tfrac{1}{2}R_{tabc}u^t. \tag{A9}$$

It can be verified that this gives for a rigid motion

$$\dot{u}_{a;b} = \dot{\omega}_{ab} - \dot{u}_a\dot{u}_b - \omega_{at}\omega_b{}^t - \ddot{u}_a u_b - \omega_{at}\dot{u}^t u_b + R_{asbt}u^s u^t, \tag{A10}$$

and therefore

$$\perp\dot{u}_{a;b} = \perp\dot{\omega}_{ab} - \dot{u}_a\dot{u}_b - \omega_{at}\omega_b{}^t + \perp R_{asbt}u^s u^t. \tag{A11}$$

In order to derive an expression for $\omega_{ab;c}$ we first observe that

$$\omega_{ab;c} = 3\omega_{[ab;c]} + 2\omega_{c[b;a]}. \tag{A12}$$

But since $R_{t[abc]} = 0$, it follows from (A9) that

$$u_{[a;bc]} = 0, \tag{A13}$$

and therefore since $u_{a;b} = \omega_{ab} - \dot{u}_a u_b$ for a rigid motion, we find that

$$\omega_{[ab;c]} = \tfrac{1}{3}(u_a \dot{u}_{[c;b]} + u_b \dot{u}_{[a;c]} + u_c \dot{u}_{[b;a]}) + \dot{u}_{[a}\omega_{bc]}. \tag{A14}$$

Also, by covariantly differentiating $\omega_{cb} = u_{c;b} + \dot{u}_c u_b$ with respect to x^a, taking the skew part on indices a and b and using (A9) it can be verified that

$$\omega_{c[b;a]} = \dot{u}_{c;[a} u_{b]} + \dot{u}_c(\omega_{ba} + \dot{u}_{[a} u_{b]}) + \tfrac{1}{2} R_{tcba} u^t. \tag{A15}$$

By substituting from (A14) and (A15) into (A12) and using (A10) we obtain

$$\omega_{ab;c} = -\omega_{ab}\dot{u}_c + 2\dot{u}_{[a}\omega_{b]c} - \dot{\omega}_{ab} u_c + 2u_{[a}\omega_{b]t}\omega_c{}^t + \perp R_{abct} u^t, \tag{A16}$$

and therefore

$$\perp\omega_{ab;c} = -\omega_{ab}\dot{u}_c + 2\dot{u}_{[a}\omega_{b]c} + \perp R_{abct} u^t. \tag{A17}$$

Using (A11) and (A17), (A8) becomes

$$h_{ar} h^s_b h^f_c h^g_d (\mathcal{L}_u \Gamma^r_{sg})_{;f} = \omega_{ac}(\omega_{bt}\omega_d{}^t + 2\dot{u}_b \dot{u}_d - \perp R_{bsdt} u^s u^t) - 2\omega_{c(b} \perp\dot{\omega}_{d)a} - 2\omega_{a(b}(\perp\dot{\omega}_{d)c} - 2\dot{u}_{d)}\dot{u}_c - \omega_{d)t}\omega_c{}^t + \perp R_{d)sct} u^s u^t) + 2\dot{u}_{(b} \perp R_{d)act} u^t. \tag{A18}$$

But for a rigid motion,

$$\perp\dot{\omega}_{ab} = \mathcal{L}_u \omega_{ab}, \tag{A19}$$

and therefore

$$2h_{ar} h^s_b h^f_c h^g_d (\mathcal{L}_u \Gamma^r_{s[g})_{;f]} = \mathcal{L}_u(2\omega_{ab}\omega_{cd} + \omega_{ac}\omega_{bd} + \omega_{ad}\omega_{cb}) + \dot{u}_b \perp R_{tacd} u^t + \dot{u}_d \perp R_{tcab} u^t + \dot{u}_c \perp R_{tdba} u^t. \tag{A20}$$

But since $\omega_{ab} = -\omega_{ba}$ and $\omega_{ab} u^b = 0$, it can be verified that

$$\omega_{a[b}\omega_{cd]} = 0 \tag{A21}$$

and therefore

$$2h_{ar} h^s_b h^f_c h^g_d (\mathcal{L}_u \Gamma^r_{s[g})_{;f]} = \mathcal{L}_u(3\omega_{ab}\omega_{cd}) + \dot{u}_b \perp R_{tacd} u^t + \dot{u}_d \perp R_{tcab} u^t + \dot{u}_c \perp R_{tdba} u^t. \tag{A22}$$

Equation (A1) follows immediately by substituting from (A5) and (A22) into (A3).

APPENDIX B: LIE DERIVATIVE IN A RIGID MOTION

We prove that if a covariant vector T_a satisfies

$$T_a u^a = 0, \quad \mathcal{L}_u T_a = 0, \tag{B1}$$

and if the motion is rigid, then

$$\mathcal{L}_u(\perp T_{a;b}) = 0. \tag{B2}$$

Consider a rigid motion. Now, using (A4) it can be shown that

$$\mathcal{L}_u(\perp T_{a;b}) = \dot{u}_a h^s_b u^r T_{r;s} + \dot{u}_b h^r_a \dot{T}_r + h^r_a h^s_b \mathcal{L}_u T_{r;s}. \tag{B3}$$

But since $u^r T_r = 0$, we have for a rigid motion,

$$h^s_b u^r T_{r;s} = \omega_{br} T^r, \tag{B4}$$

and since $\mathcal{L}_u T_a = 0$ it follows for a rigid motion that

$$h^r_a \dot{T}_r = \omega_{ar} T^r. \tag{B5}$$

Also, in general,[14]

$$\mathcal{L}_u(T_{r;s}) = (\mathcal{L}_u T_r)_{;s} - T_c \mathcal{L}_u \Gamma^c_{rs} \tag{B6}$$

and therefore since $\mathcal{L}_u T_r = 0$ we have

$$\mathcal{L}_u(T_{r;s}) = -T_c \mathcal{L}_u \Gamma^c_{rs}. \tag{B7}$$

Hence by substituting from (B4), (B5), and (B7) into (B3) we obtain

$$\mathcal{L}_u(\perp T_{a;b}) = 2\dot{u}_{(a}\omega_{b)c} T^c - h^r_a h^s_b (\mathcal{L}_u \Gamma^c_{rs}) T_c. \tag{B8}$$

But by using (A6) and (A7) a direct calculation gives, for a rigid motion,

$$h^r_a h^s_b (\mathcal{L}_u \Gamma^c_{rs}) T_c = 2\dot{u}_{(a}\omega_{b)c} T^c. \tag{B9}$$

Equation (B2) follows immediately from (B8) and (B9).

APPENDIX C: ISOMETRIC MOTION

We prove that if the motion is isometric then

$$\mathcal{L}_u E_{ab} = 0, \tag{C1}$$

$$\mathcal{L}_u H_{ab} = 0, \tag{C2}$$

$$\mathcal{L}_u(\perp G_{ab} - \tfrac{1}{3}(h^{st} G_{st}) h_{ab}) = 0. \tag{C3}$$

Consider an isometric motion. Then there exists a Killing vector field $\xi^a = \xi u^a$, $\xi = (-\xi_a \xi^a)^{1/2}$, such that

$$\mathcal{L}_\xi g_{ab} = 0. \tag{C4}$$

By contracting (C4) in turn with $u^a u^b$ and with u^b it can be shown that

$$\dot{\xi} = 0, \quad \xi_{,a} = \xi \dot{u}_a, \tag{C5}$$

and therefore that

$$\mathcal{L}_\xi u^a = 0, \quad \mathcal{L}_\xi h^{ab} = 0, \quad \mathcal{L}_\xi h^a_b = 0, \quad \mathcal{L}_\xi h_{ab} = 0, \tag{C6}$$

results which will be used below. The operation of raising or lowering an index commutes with the Lie derivative along ξ^a and therefore for instance

$$\mathcal{L}_\xi T_{ab\cdots c} = 0 \Leftrightarrow \mathcal{L}_\xi T^a{}_{b\cdots c} = 0. \tag{C7}$$

Also it is readily verified that if $T_{ab\cdots c}$ is orthogonal to u^a on all indices then

$$\mathcal{L}_u T_{ab\cdots c} = (1/\xi)\mathcal{L}_\xi T_{ab\cdots c}. \tag{C8}$$

Since for any vector ξ^a (Ref. 14)

$$\mathcal{L}_\xi R^a{}_{bcd} = 2(\mathcal{L}_\xi \Gamma^a_{b[d})_{;c]} \tag{C9}$$

and

$$\mathcal{L}_\xi \Gamma^a_{bc} = \tfrac{1}{2} g^{at}[(\mathcal{L}_\xi g_{bt})_{;c} + (\mathcal{L}_\xi g_{ct})_{;b} - (\mathcal{L}_\xi g_{bc})_{;t}] \tag{C10}$$

it follows for a Killing vector ξ^a that

$$\mathcal{L}_\xi R^a{}_{bcd} = 0. \tag{C11}$$

Hence

$$\mathscr{L}_\xi R_{ab} = 0, \quad \mathscr{L}_\xi R = 0, \quad \mathscr{L}_\xi G_{ab} = 0, \tag{C12}$$

and therefore by Eq. (3.7)

$$\mathscr{L}_\xi C_{abcd} = 0. \tag{C13}$$

Thus by (C6) and (C13),

$$\mathscr{L}_\xi E_{ab} = \mathscr{L}_\xi (C_{arbs} u^r u^s) = 0, \tag{C14}$$

and since $E_{ab} u^a = 0 = E_{ab} u^b$ it follows from (C8) that $\mathscr{L}_u E_{ab} = 0$. Also

$$\mathscr{L}_\xi \eta_{abcd} = \eta_{abcd} \xi^t{}_{;t} = 0 \tag{C15}$$

since $\xi^t{}_{;t} = 0$ for a Killing vector, and therefore

$$\mathscr{L}_\xi H_{ab} = \mathscr{L}_\xi (\tfrac{1}{2}\eta_{arpq} C^{pq}{}_{bs} u^r u^s) = 0. \tag{C16}$$

Since $H_{ab} u^a = 0 = H_{ab} u^b$ it follows from (C8) that $\mathscr{L}_u H_{ab} = 0$. Finally, it is readily verified using (C6) and (C12) that

$$\mathscr{L}_\xi (\perp G_{ab} - \tfrac{1}{3}(h^{st} G_{st}) h_{ab}) = 0, \tag{C17}$$

and since $\perp G_{ab} - \tfrac{1}{3}(h^{st} G_{st}) h_{ab}$ is orthogonal to u^a on both indices, (C3) follows directly from (C8).

[1]M. Born, Ann. Phys. (Leipzig) **30**, 1 (1909).
[2]G. Herglotz, Ann. Phys. (Leipzig) **31**, 393 (1910).
[3]F. Noether, Ann. Phys. (Leipzig) **31**, 919 (1910).
[4]G. Salzman and A. H. Taub, Phys. Rev. **95**, 1659 (1954).
[5]C. B. Rayner, C. R. Acad. Sci. **248**, 929 (1959).
[6]F. A. E. Pirani and G. Williams, Séminaire Janet (Mécanique analytique et Mécanique céleste), Vol. 5, No. 8 (1962).
[7]A. Trautman, in *Lectures in General Relativity*, Brandeis Summer Institute in Theoretical Physics, edited by S. Deser and K. Ford (Prentice-Hall, Englewood Cliffs, NJ, 1964), Vol. 1, Chap. 8, pp. 201–227.
[8]J. L. Synge, *Relativity: The Special Theory* (North-Holland, Amsterdam, 1956), Chap. VIII; *Relativity: The General Theory* (North-Holland, Amsterdam, 1971), Chap. IV.
[9]G. Williams, J. Math. Anal. Appl. **22**, 646 (1968).
[10]R. H. Boyer, Proc. R. Soc. London Ser. A **283**, 343 (1965).
[11]G. F. R. Ellis, in *General Relativity and Gravitation*, edited by R. K. Sachs (Academic, New York, 1971), pp. 104–182.
[12]G. F. R. Ellis, in *Cargèse Lectures in Physics*, edited by E. Schatzman (Gordon and Breach, New York, 1973), Vol. 6, pp. 1–60.

[13]Latin indices run over the four coordinates of space-time, Greek indices over the three spatial coordinates. A semicolon denotes covariant differentiation with respect to the metric tensor g_{ab} of space-time [signature $(-+++)$]. An overhead dot denotes covariant differentiation along a particle world line; for example,

$$\dot{A}^a = A^a{}_{;b} u^b.$$

The Reimann curvature tensor is defined through the identity

$$A_{a;[bc]} = 2R_{tabc} A^t$$

and the Ricci tensor R_{ab}, the Ricci scalar R, and the Einstein tensor G_{ab}, are defined, respectively, by

$$R_{ab} = R^t{}_{atb}, \quad R = R^a{}_a, \quad G_{ab} = R_{ab} - \tfrac{1}{2} R g_{ab}.$$

Units are used in which the speed of light in vacuum and Einstein's gravitational constant are both unity. Einstein's field equations are

$$G_{ab} + \Lambda g_{ab} = T_{ab}.$$

The projection symbol $\perp$ before a tensor expression denotes that after all other indicated contractions in that expression have been performed, each remaining free index is projected with h^a_b, where $h^a_b = g^a_b + u^a u_b$ is the projection tensor onto the three-plane orthogonal to u^a; for example,

$$\perp R^a{}_{bcd} = h^a_r h^s_b h^t_c h^p_d R^r{}_{stp},$$

$$\perp R_{abcd} u^d = h^r_a h^s_b h^t_c R_{rstd} u^d.$$

[14]K. Yano, *The Theory of Lie Derivatives and its Applications* (North-Holland, Amsterdam, 1955), pp. 16, 17, and 52.
[15]J. Ehlers, in *Relativity, Astrophysics and Cosmology*, edited by W. Israel (Reidel, Dordrecht, 1972), pp. 89–98.

[16]η^{abcd} is defined by

$$\eta^{abcd} = \eta^{[abcd]}, \quad \eta^{0123} = (-g)^{-1/2},$$

where $g = \det[g_{ab}]$. We make use of the following identities:

$$\eta^{abcd}\eta_{efgh} = -4!\delta^{[a}_e \delta^b_f \delta^c_g \delta^{d]}_h,$$

$$\eta^{abcd}\eta_{afgh} = -3!\delta^{[b}_f \delta^c_g \delta^{d]}_h,$$

$$\eta^{abcd}\eta_{abgh} = -4\delta^{[c}_g \delta^{d]}_h.$$

[17]The tensor $\perp G_{ab} - \tfrac{1}{3}(h^{st} G_{st}) h_{ab}$ can be expressed in terms of R_{ab} instead of G_{ab}; we have

$$\perp G_{ab} - \tfrac{1}{3}(h^{st} G_{st}) h_{ab} = \perp R_{ab} - \tfrac{1}{3}(h^{st} R_{st}) h_{ab}.$$

A class of solutions for self-gravitating dust in Newtonian gravity

Thomas Buchert and Günter Götz
Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, D-8046 Garching b. München, West Germany

(Received 30 September 1986; accepted for publication 27 May 1987)

The Lagrange method is used to obtain a class of solutions of the three-dimensional hydrodynamical equations governing the motion of matter with vanishing pressure in its own Newtonian gravitational field. The class is characterized by the property that each fluid particle has constant acceleration. The class contains rotational and irrotational flows. For rotational flows the expansion tensor has one zero eigenvalue, while for irrotational flows it has two zero eigenvalues, which implies that every fluid element contracts or expands in two or one spatial directions, respectively; nevertheless, the density depends on all three coordinates. The general one-dimensional solution is included as a subclass.

I. INTRODUCTION

The study of solutions of the equations of a hydrodynamical system including self-gravity has been, on one hand, the subject of astrophysicists focusing on spherically symmetric stellar collapse, and on the other hand, the subject of cosmologists interested in the process of galaxy formation dominated by nonspherical motions. The exact solutions including pressure are, for spherical symmetry, based on polytropic equations of state providing stellar models (time-dependent models are treated by Munier and co-workers,[1] Glass,[2] and others); for dust, the evolution of spherically symmetric inhomogeneities is studied, e.g., by Henriksen and Robertis.[3] Spatially homogeneous solutions are discussed in Heckmann and Schücking.[4] The dynamics of homogeneous ellipsoids are also treated in a paper by Barrow and Silk.[5] Up to now almost nothing is known about inhomogeneous, anisotropic solutions (for the dust case, similarity solutions have been studied by Fillmore and Goldreich[6] for spherical, cylindrical, and planar symmetry); the Newtonian analogs of Szekeres space-times in general relativity have been discussed by Lawitzky.[7]

This work is primarily motivated by the intention to get a precise understanding of the role played by nonlinearities in the evolution of density inhomogeneities. In the present paper we concentrate on the zero pressure case and obtain solutions that remain valid until singularities in the density develop, i.e., in the regime where perturbation theories fail.

One fruitful effort to go beyond perturbation analysis is the so-called "pancake-scenario" in galaxy formation theory conceived by Zel'dovich[8] in the 1970's. He proposed an "approximate nonlinear theory" based on assumptions extrapolating the linear theory of gravitational instability. This scenario contains key elements relevant to the evolution of large scale structure in the universe (see the review by Shandarin *et al.*[9]). The latter is ascribed to the caustic structure of the underlying evolution ansatz (Arnol'd, Shandarin, and Zel'dovich[10]). A realization of the "pancake-scenario" is achieved, e.g., by the assumption that "cold dark matter" (i.e., dust) is the dominant matter constituent. Another approach to tackle the problem of nonlinear evolution uses numerical simulations (Davies *et al.*[11] and Shapiro *et al.*[12]; for spherical symmetry a survey by Woodward[13]). However, in view of the high numerical resolution that would be required to detect certain nonlinear phenomena such as solitons or self-focusing singularities (Kates[14]), it is important to pursue analytic studies. This paper is an attempt to establish the basis for the treatment of more complex problems, e.g., the inclusion of pressure and viscosity.

The article is organized as follows: Section II presents the basic equations in Eulerian and Lagrangian form, Sec. III contains the solution obtained through an ansatz suggested by the one-dimensional case, and a discussion of the solution is left to Sec. IV.

II. THE SYSTEM OF EQUATIONS

Denoting the density, the velocity, and the acceleration field with respect to some nonrotating frame (Eulerian description) by $\rho(\mathbf{x},t)$, $\mathbf{v}(\mathbf{x},t)$, and $\mathbf{g}(\mathbf{x},t)$, respectively, the equations for self-gravitating dust are

$$\rho_t + \nabla\cdot(\rho\mathbf{v}) = 0, \tag{1a}$$

$$\mathbf{v}_t + (\mathbf{v}\cdot\nabla)\mathbf{v} = \mathbf{g}, \tag{1b}$$

$$\nabla\times\mathbf{g} = 0 \Rightarrow \mathbf{g} =: -\nabla\phi, \tag{1c}$$

$$\nabla\cdot\mathbf{g} = -4\pi G\rho, \tag{1d}$$

$$\rho \geqslant 0. \tag{1e}$$

Equations (1c) and (1d) are equivalent to Poisson's equation for the gravitational potential ϕ. These equations retain their form if transformed to a translationally accelerated frame provided ϕ is appropriately transformed.[4]

To obtain solutions it is useful to transform from Eulerian coordinates x_i to Lagrangian coordinates X_i ($i = 1,2,3$) leaving the time unchanged. Lagrangian coordinates are defined as the labels $\mathbf{X}$ of the integral curves $\mathbf{x} = \mathbf{f}(\mathbf{X},t)$ of the vector field $\mathbf{v}(\mathbf{x},t)$,[15]

$$\frac{d\mathbf{f}}{dt} = \mathbf{v}(\mathbf{f},t), \quad \mathbf{f}(\mathbf{X},0) = \mathring{\mathbf{f}}(\mathbf{X}) = \mathbf{X}, \tag{2}$$

which are interpreted as the paths followed by fluid particles initially at the point $\mathbf{f}(\mathbf{X},0)$. Recall that for any vector function $\mathbf{F}(\mathbf{x},t)$ one has

$$\dot{\mathbf{F}}(\mathbf{x},t)\,|_{\mathbf{X}=\text{const}} = \dot{\mathbf{F}}(\mathbf{f}(\mathbf{X},t),t) = \partial_t\mathbf{F}(\mathbf{X},t). \tag{3}$$

 0022-2488/87/112714-06$02.50

Now $\mathbf{x}$ becomes a dependent variable and is considered as a vector field in the Lagrangian picture, which gives the transformation back to the Eulerian picture and determines completely the dynamics of the fluid as long as the inverse transformation exists, i.e., the Jacobian remains nonsingular:

$$J_{ik}(\mathbf{X},t) := \frac{\partial f_i}{\partial X_k}, \quad \det(J_{ik}) =: J \neq 0. \tag{4a}$$

Recall also that the following evolution equations for the Jacobian and its determinant hold[16]:

$$\dot{J}_{ik} = \frac{\partial v_i}{\partial X_k} = \frac{\partial v_i}{\partial x_j} J_{jk}, \quad \dot{J} = \nabla\cdot\mathbf{v}\, J. \tag{4b}$$

Upon integrating the continuity equation (1a) with the help of (4b) and expressing the operator ∇ in terms of derivatives with respect to X_i, the system (1) is cast into the following form:

$$\rho = \rho_0(\mathbf{X}) J^{-1}, \tag{5a}$$

$$\mathbf{g} = \dot{\mathbf{v}} = \ddot{\mathbf{f}}, \tag{5b}$$

$$g_{i,k} J^{-1}_{kj} = g_{j,k} J^{-1}_{ki} \quad (i \neq j), \tag{5c}$$

$$g_{i,k} J^{-1}_{ki} = -4\pi G \rho. \tag{5d}$$

Inserting (5a) and (5b) into (5c) and (5d), we obtain a system of four equations for the Jacobian:

$$\ddot{J}_{ik} J^{-1}_{kj} = \ddot{J}_{jk} J^{-1}_{ki} \quad (i \neq j), \tag{6a}$$

$$\ddot{J}_{ik} J^{-1}_{ki} = -4\pi G \rho_0(\mathbf{X}) J^{-1}. \tag{6b}$$

Denoting the functional determinant of any three functions $h_1(\mathbf{X},t)$, $h_2(\mathbf{X},t)$, and $h_3(\mathbf{X},t)$ by

$$D(h_1,h_2,h_3) := \frac{\partial(h_1,h_2,h_3)}{\partial(X_1,X_2,X_3)},$$

the system (6) may be finally cast into the form that provides the basis for seeking solutions to the present problem:

$$\epsilon_{pq[j} D(\ddot{f}_{i]}, f_p, f_q) = 0 \quad \text{(three equations)}, \tag{7a}$$

$$\epsilon_{abc} D(\ddot{f}_a, f_b, f_c) = -8\pi G \rho_0 \quad \text{(one equation)}. \tag{7b}$$

(Note: $J^{-1}_{ab} = \mathrm{ad}(J_{ab}) J^{-1} = [1/(2J)] \epsilon_{ajk} \epsilon_{blm} J_{lj} J_{mk}$.)

III. THE SOLUTION

We consider the class of motions characterized by the property that fluid particles be uniformly accelerated along their paths, labeled by $\mathbf{X}$, i.e.,

$$\mathbf{g}(\mathbf{X},t) = \mathbf{g}(\mathbf{X},0) =: \mathring{\mathbf{g}}(\mathbf{X}). \tag{8a}$$

This ansatz implies, by means of Eq. (5b),

$$\mathbf{v}(\mathbf{X},t) = \mathring{\mathbf{v}}(\mathbf{X}) t + \mathring{\mathbf{g}}(\mathbf{X}) t, \tag{8b}$$

$$\mathbf{f}(\mathbf{X},t) = \mathbf{X} + \mathring{\mathbf{v}}(\mathbf{X}) t + \mathring{\mathbf{g}}(\mathbf{X}) t^2/2. \tag{8c}$$

The derivative of the map $\mathbf{X} \to \mathbf{x}$ at time t is, according to (8c),

$$J_{ik}(\mathbf{X},t) = \delta_{ik} + \mathring{v}_{i,k} t + \mathring{g}_{i,k} t^2/2 \tag{8d}$$

[$\mathring{\mathbf{g}}(\mathbf{X})$, $\mathring{\mathbf{v}}(\mathbf{X})$, and $\mathring{\mathbf{f}}(\mathbf{X},0) = \mathbf{X}$ are the initial data of the fields $\mathbf{g}$, $\mathbf{v}$, and $\mathbf{f}$, given at the instant $t = 0$].

Uniform acceleration is a general property of one-dimensional flows, as can be seen from Eqs. (1a) and (1d) restricted to one spatial dimension:

$$\rho_t + (\rho v)_x = 0, \quad g_x = -4\pi G \rho.$$

Inserting the second equation into the first we find $\dot{g} = \beta(t)$. Without loss of generality α may be set equal to zero since β is only associated with a spatially constant velocity field and does not alter the density distribution (the basic equations are "translation covariant," see Heckmann and Schücking[4b]); i.e., the ansatz (8a) yields the general solution in this case. [Note that the field equation (1c) is trivially satisfied in one dimension.]

In the three-dimensional case we proceed as follows: We plug (8c) into the system of equations (7) whose left-hand sides yield polynomials in t whose coefficients must vanish. The resulting 17 equations, which constrain the coefficient functions $\mathring{\mathbf{v}}(\mathbf{X})$ and $\mathring{\mathbf{g}}(\mathbf{X})$, are given in Appendix A.

Lemma: Relations (A2)–(A4) and (A6)–(A9) listed in Appendix A are satisfied if and only if the six functions $\mathring{v}_1$, $\mathring{v}_2$, $\mathring{v}_3$ and $\mathring{g}_1$, $\mathring{g}_2$, $\mathring{g}_3$ depend functionally on a single function only, which—without loss of generality—is taken as $\mathring{g}_1$:

$$\mathring{v}_1 = F(\mathring{g}_1), \tag{9a}$$

$$\mathring{v}_2 = H(\mathring{g}_1), \tag{9b}$$

$$\mathring{v}_3 = K(\mathring{g}_1), \tag{9c}$$

$$\mathring{g}_2 = A(\mathring{g}_1), \tag{9d}$$

$$\mathring{g}_3 = B(\mathring{g}_1). \tag{9e}$$

Relation (A1) implies that

$$\mathring{g}_1 = E(X_1 + A'X_2 + B'X_3), \tag{9f}$$

where E is arbitrary, (A5) simply determines ρ_0. (The proof is given in Appendix A.)

This formulation of the solutions of Eqs. (A2)–(A4) and (A6)–(A9) avoids using explicitly the gravitational potential $\mathring{\phi}$, whose existence is guaranteed by 1(c). This $\mathring{\phi}$ is defined by

$$\mathring{\phi}(\mathbf{X}) := \phi(\mathbf{X},0), \quad \mathbf{g}(\mathbf{X},t) \overset{(8a)}{=} -\nabla\phi(\mathbf{X},t) = -\nabla\phi|_{t=0} = -\nabla_{\mathbf{X}}\mathring{\phi}. \tag{10}$$

An alternative formulation of the Lemma provides the following corollary.

Corollary: Taking the function on which all the others have to depend as $\mathring{\phi}_{,1}$, the complete set of restrictions (A1)–(A9) reads

$$\mathring{\phi}_{,2,2}\mathring{\phi}_{,3,3} - \mathring{\phi}^2_{,2,3} = 0, \tag{11a}$$

$$\mathring{\phi}_{,3,3}\mathring{\phi}_{,1,1} - \mathring{\phi}^2_{,3,1} = 0, \tag{11b}$$

$$\mathring{\phi}_{,1,1}\mathring{\phi}_{,2,2} - \mathring{\phi}^2_{,1,2} = 0, \tag{11c}$$

$$\mathring{v}_1 = F(\mathring{\phi}_{,1}), \tag{11d}$$

$$\mathring{v}_2 = H(\mathring{\phi}_{,1}), \tag{11e}$$

$$\mathring{v}_3 = K(\mathring{\phi}_{,1}). \tag{11f}$$

One of the equations [(11a)–(11c)] is redundant. The general local solution of the restrictions[17] (11a)–(11c) can be written down in terms of three functions A,B,C (see Ref. 18):

$$\mathring{\phi} = \alpha X_1 + A(\alpha) X_2 + B(\alpha) X_3 + C(\alpha), \tag{12a}$$

$$0 = X_1 + A'(\alpha) X_2 + B'(\alpha) X_3 + C'(\alpha) \tag{12b}$$

(where α is written for $\mathring{\phi}_{,1}$ and the prime denotes a derivative with respect to the argument). This implies that the derivatives of $\mathring{\phi}$ are functionally dependent

$$[\mathring{\phi}_{,1} = \alpha, \mathring{\phi}_{,2} = A(\alpha), \mathring{\phi}_{,3} = B(\alpha)].$$

The functions A and B are the same as those we used in (9), the function C is related to the inverse function of E by $-C'(\alpha) = E^{-1}(\alpha)$. [The proof that all local solutions of (11a)–(11c) are determined by (12) is given in Appendix B.]

Equations (11a)–(11c) express the property that the two-surfaces $\phi(X_i = \text{const}\, X_j, X_k)$ $(i,j,k = 1,2,3$ pairwise different) have zero Gaussian curvature everywhere, i.e., have to be developable surfaces.

Theorem 1: The solution of Eqs. (1) under assumption (8a) is given by

$$\rho(\mathbf{X},t) = \frac{\rho_0}{1 + (\mathring{v}_{1,1} + \mathring{v}_{2,2} + \mathring{v}_{3,3})t - 4\pi G\rho_0 t^2/2}, \quad \rho_0 \geqslant 0, \tag{13a}$$

$$\rho(\mathbf{X},0) = \rho_0(\mathbf{X}) = \Delta_{\mathbf{X}}\mathring{\phi}/4\pi G = -\frac{(1 + A'^2 + B'^2)}{4\pi G(A''X_2 + B''X_3 + C'')},$$

$$\mathbf{v}(\mathbf{X},t) = \mathring{\mathbf{v}} - \nabla_{\mathbf{X}}\mathring{\phi}t, \qquad \mathbf{v}(\mathbf{X},0) = \mathring{\mathbf{v}}(\mathbf{X}), \tag{13b}$$

$$f(\mathbf{X},t) = \mathbf{X} + \mathring{\mathbf{v}}t - \nabla_{\mathbf{X}}\mathring{\phi}t^2/2, \qquad f(\mathbf{X},0) = \mathbf{X}, \tag{13c}$$

where $\mathring{\mathbf{v}}$ and $\mathring{\phi}$ are constrained by (11). The restrictions on $\mathring{\phi}$ are equivalent to the restrictions (9) on $\mathring{\mathbf{g}}$. [The proof of the theorem follows from the Lemma and the Corollary by means of (5a), (8b), and (8c).]

Note that the determinant of the Jacobian (8d) reads

$$J = \tfrac{1}{6}\delta_{ab}\epsilon_{ijk}\epsilon_{bcd}J_{ai}J_{cj}J_{dk},$$

which reduces to

$$J = 1 + \text{Tr}(\mathring{v}_{i,k})t + \text{Tr}(\mathring{g}_{i,k})t^2/2,$$

in view of the constraints that have to be imposed on the initial data (11).

A particular solution is obtained by prescribing A,B,C and F,H,K, solving (12b) for α which in turn yields $\mathring{\phi}$ via (12a). Elimination of $\mathbf{X}$ inverts the transformation from Eulerian to Lagrangian coordinates and yields the Eulerian description of the solutions $\rho(\mathbf{x},t)$ and $\mathbf{v}(\mathbf{x},t)$. The density given in (13a) has to be positive for the whole range of $\mathbf{X}$ if one is interested in physically meaningful solutions. This requires a choice of the functions A,B,C and such that

$$A''X_2 + B''X_3 + C'' < 0.$$

This is possible in the one-dimensional case. Thus the three-dimensional solutions can have this property at least in some small neighborhood of the submanifold in (A,B,C) space, which generates the plane-symmetric solutions with $\rho_0 > 0$.

Finally we remark that the class of solutions presented above contains the general one-dimensional solution as a subclass: Assuming $\mathring{v}_2 = \mathring{v}_3 = 0$ and vanishing derivatives with respect to X_2 and X_3 shows that Eqs. (11a)–(11f) no longer provide restrictions on the initial data [$\mathring{v}_1 = F(\mathring{\phi}_{,1})$ is locally true for functions of one variable]. The only nontrivial field equation (5d) reads

$$\frac{\partial g(\mathbf{X},t)}{\partial X}\frac{\partial X}{\partial x} = \frac{\partial \mathring{g}(\mathbf{X})}{\partial X}\frac{\partial X}{\partial x}$$

and is solved uniquely by (8a); the solution is general since one can choose $\rho_0(X)$ and the initial velocity field $\mathring{v}_1(X)$ arbitrarily.

IV. DISCUSSION

In this section we focus on the kinematical properties of the solution, which helps us to understand the kinds of motions admitted. For this purpose we adopt the standard decomposition of the velocity gradient into its symmetric and antisymmetric parts:

$$\frac{\partial v_i}{\partial x_j} = \omega_{ij} + \theta_{ij} = \omega_{ij} + \sigma_{ij} + \frac{1}{3}\theta\delta_{ij}, \tag{14}$$

where $\omega_{ij} = \partial v_{[i}/\partial x_{j]}$ represents the vorticity tensor and $\theta_{ij} = \partial v_{(i}/\partial x_{j)}$ the expansion tensor, split into its trace-free part σ_{ij}, which represents the shear of the flow, and its trace θ, the expansion rate.

Let us now concentrate on single particle paths. For the class of solutions presented, the kinematical variables along these paths are, in terms of initial conditions,

$$\omega_{ik}(\mathbf{X},t) = \tfrac{1}{2}(\mathring{v}_{i,k} - \mathring{v}_{k,i})J^{-1}, \tag{15}$$

$$\theta_{ik}(\mathbf{X},t) = (\tfrac{1}{2}(\mathring{v}_{i,k} + \mathring{v}_{k,i}) + \mathring{g}_{i,k}t)J^{-1}, \tag{16a}$$

$$\sigma_{ik}(\mathbf{X},t) = \theta_{ik}(\mathbf{X},t) - \tfrac{1}{3}(\mathring{v}_{i,i} - 4\pi G\rho_0 t)\delta_{ik}J^{-1}, \tag{16b}$$

$$\theta(\mathbf{X},t) = (\mathring{v}_{i,i} - 4\pi G\rho_0 t)J^{-1}. \tag{16c}$$

The scalar θ measures the rate of change of a fluid element's volume. Equation (16c) shows that, for the time interval $0 \leqslant t < t_c$ $[J(t_c) = 0]$, θ lies within the bounds $v_{i,i} \geqslant \theta > -\infty$, if $\rho_0 > 0$, which shows that, irrespective of (sufficiently smooth) initial conditions, the collapse cannot be avoided ($t \to t_c$, $\theta \to -\infty$). If the initial divergence of the velocity field is positive, then θ is positive (expansion) during the time $\Delta t = v_{i,i}/4\pi G\rho_0$. The shear scalar σ, defined by $\sigma^2 := \tfrac{1}{2}\sigma_{ik}\sigma_{ki}$, which measures the anisotropy of the motion, is

$$\sigma^2(\mathbf{X},t) = \tfrac{1}{4}((v_{1,1})^2 + (v_{2,2})^2 + (v_{3,3})^2)J^{-2} + \omega_i\omega_i/2, \tag{17}$$

where $\omega_i = \tfrac{1}{2}\epsilon_{ijk}\omega_{jk}$ are the components of the rotation vector $\boldsymbol{\omega} = \tfrac{1}{2}\,\text{rot}\,\mathbf{v}$. Starting out with initial isotropy $[\sigma^2(X,0) = 0]$, which is only possible for $v_{i,k} = 0$, the motion always becomes anisotropic for $t > 0$.

A refined measure for the anisotropy of the expansion/contraction of fluid elements is provided by an investigation of the eigenvalues of the expansion tensor. These eigenvalues are the extrema of the expansion rate along the directions of the principle axes of θ_{ik}. They are the solutions of the characteristic polynomial

$$\lambda^3 - \lambda^2\,\text{I} + \lambda\,\text{II} - \text{III} = 0, \tag{18}$$

where

$$\text{I} := \theta,$$

$$\text{II} := \theta_{22}\theta_{33} + \theta_{33}\theta_{11} + \theta_{11}\theta_{22} - \theta_{23}^2 - \theta_{31}^2 - \theta_{12}^2,$$

$$\text{III} := \det(\theta_{ik})$$

denote the three scalar invariants of θ_{ik}.

For our solution it turns out that[19]

$$\mathrm{II} = -\omega_i\omega_i \leqslant 0, \quad \mathrm{III} = 0, \tag{19a}$$

which implies

$$\lambda_1 = \tfrac{1}{2}(\theta + \sqrt{\theta^2 - 4\mathrm{II}}) > 0,$$
$$\lambda_2 = 0, \tag{19b}$$
$$\lambda_3 = \tfrac{1}{2}(\theta - \sqrt{\theta^2 - 4\mathrm{II}}) < 0.$$

Nonvanishing rotation causes the fluid elements to expand in one direction and to contract in another for all times. The rotation vector $\boldsymbol{\omega}$ is an eigenvector of the expansion tensor with eigenvalue $\lambda_2 = 0$ ($\theta_{ik}\omega_k = 0$). Equation (15) shows that the direction of $\boldsymbol{\omega}$ is locally preserved along the particle path, the modulus of $\boldsymbol{\omega}$ increasing proportionally to the density. This implies that $\boldsymbol{\omega}/\rho$ is a conserved quantity of our solution.

For potential flows, the expansion tensor $\theta_{ik} = \partial v_i/\partial x_k$ ($\omega_{ik} = 0$) has only one nonvanishing eigenvalue [see (19)], meaning that every fluid element expands/contracts along one spatial direction. Note that these directions are, in general, different for different paths and vary smoothly from point to point. An immediate consequence of this behavior of fluid elements is the tendency for the matter to pile up into flat structures; volume elements located at $\mathbf{X}$ degenerate into elements of caustics in Eulerian space at times $t_c(\mathbf{X})$. The caustics (surfaces of infinite density) are given by the image of the singular set $\$$ of points in Lagrangian space at a time t' [see Eq. (13a)]:

$$\$ = \{(\mathbf{X},t')/J(\mathbf{X},t') = 1 + \nabla_{\mathbf{X}}\cdot\mathring{\mathbf{v}}t' - 4\pi G\rho_0 t'^2/2 = 0\}. \tag{20}$$

Vanishing rotation implies restrictions on the functions F, H, and K [Eqs. (11d)–(11f)], which, expressed in terms of the velocity potential $\mathring{S}(\mathbf{X})$ ($\mathbf{v} = \nabla S|_{t=0} = \nabla_{\mathbf{X}}\mathring{S}$), are

$$\mathring{S}_{,1} = F(\mathring{\phi}_{,1}), \tag{21a}$$
$$\mathring{S}_{,2} = H(\mathring{\phi}_{,1}), \tag{21b}$$
$$\mathring{S}_{,3} = K(\mathring{\phi}_{,1}), \tag{21c}$$
$$F'A' = H', \tag{21d}$$
$$F'B' = K'. \tag{21e}$$

To obtain Eqs. (21d) and (21e) we used the potential character to relate F, H, and K via Eqs. (21a)–(21c) and expressed these relations in terms of A and B by means of (12). According to the Kelvin–Helmholtz circulation theorem, the potential character of the flow is preserved in time. From Eqs. (21a)–(21c) follow restrictions completely equivalent to those imposed on $\mathring{\phi}$ by Eqs. (11a)–(11c):

$$\mathring{S}_{,2,2}\mathring{S}_{,3,3} - \mathring{S}^2_{,2,3} = 0, \tag{22a}$$
$$\mathring{S}_{,3,3}\mathring{S}_{,1,1} - \mathring{S}^2_{,3,1} = 0, \tag{22b}$$
$$\mathring{S}_{,1,1}\mathring{S}_{,2,2} - \mathring{S}^2_{,1,2} = 0. \tag{22c}$$

(Again, one of these equations is redundant.)

Expressing the solution for the potential $\mathring{S}$ in a manner similar to Eqs. (12) in terms of the parameter $\alpha = \mathring{\phi}_{,1}$, we have

$$\mathring{S} = F(\alpha)X_1 + H(\alpha)X_2 + K(\alpha)X_3 + \widetilde{C}(\alpha), \tag{23a}$$

$$0 = F'(\alpha)X_1 + H'(\alpha)X_2 + K'(\alpha)X_3 + \widetilde{C}'(\alpha)$$
$$(\widetilde{C}' = F'C'). \tag{23b}$$

We can choose, instead of six, only four functions independently. The choice $F' = 1$, i.e., F is the identity, means that $\mathring{\phi}$ and $\mathring{S}$ are identical surfaces. This implies that the complete set[17] of functions Ω for which the two-surfaces $\Omega(X_i = \text{const}, X_j, X_k)$ ($i,j,k = 1,2,3, i \neq k$) have vanishing Gaussian curvature generate the solution.

One may ask whether there are further solutions that have the property of two vanishing eigenvalues of the expansion tensor.

Theorem 2: In the class of all irrotational flows our solution is the general three-dimensional one with the property of two vanishing eigenvalues of the expansion tensor.

Proof: We consider the general equations (6) and follow the motion of a fluid element along some fixed path γ: $t \to \mathbf{f}(\mathbf{X}',t)$. We fix a coordinate system at $t = 0$ such that the directions of the coordinate axes coincide with the eigendirections of the expansion tensor. Along a path γ, these axes do not change in time, since the transport equation for any vector n along the path is (Ellis[20])

$$\delta_{jk}\dot{n}_k = (\omega_{jk} + \sigma_{jk} - (\sigma_{ab}n_a n_b)\delta_{jk})n_k.$$

If $\sigma_{jk}n_k \propto n_j$ and $\omega_{jk} = 0$, then $\delta_{jk}\dot{n}_k = 0$.

The fact that θ_{ij} remains diagonal renders the Jacobian diagonal along the path [see Eq. (4b)]. Now, Eq. (6a) is identically satisfied and Eq. (6b) reads

$$\ddot{J}_{11}J_{22}J_{33} + \ddot{J}_{22}J_{33}J_{11} + \ddot{J}_{33}J_{11}J_{22}|_{\mathbf{X}=\mathbf{X}'} = -4\pi G\rho_0. \tag{24}$$

If we assume two eigenvalues of θ_{ij} to be zero,

$$\theta_{22}(\mathbf{X}=\mathbf{X}',t) = \theta_{33}(\mathbf{X}=\mathbf{X}',t) = 0,$$

then from (24) it follows that $\dddot{J}_{11} = \dot{g}_{1,1} = 0$, which yields exactly the assumption (8a) on which our solution is based. This argument holds for any γ.

This shows that the solution is an exact description for the idealized situation where fluid elements contract in one preferred spatial direction (which is different for different fluid elements) and where the motion in the other two directions is absent. Therefore the solution might reflect general properties of highly anisotropic motions.

Further papers are in preparation, one of them concerning solutions for the dynamics of dust relative to an isotropically expanding homogeneous background; it will be shown there that the class of solutions, which corresponds to the class considered here, contains the self-consistent[21] part of the "approximate ansatz" of "pancake-theory" proposed by Zel'dovich[8] as a subclass. Also solutions for self-gravitating fluids including pressure will be given; it will be discussed that solitons may arise in such a system.

ACKNOWLEDGMENTS

We like to thank Ron Kates and Jürgen Ehlers for helpful discussions.

APPENDIX A: RESTRICTIONS ON INITIAL DATA: (PROOF OF THEOREM 1)

Inserting our ansatz [Eq. (8)] into the system of equations (7) we obtain explicitly from (7a)

$$t^0: D(\mathring{g}_k,\mathring{f}_k,\mathring{f}_i) = D(\mathring{g}_j,\mathring{f}_i,\mathring{f}_j) \Leftrightarrow \nabla_X \times \mathring{\mathbf{g}} = 0, \tag{A1}$$

$$t^1: D(\mathring{g}_k,\mathring{v}_k,\mathring{f}_i) + D(\mathring{g}_k,\mathring{f}_k,\mathring{v}_i) = D(\mathring{g}_j,\mathring{f}_i,\mathring{v}_j) + D(\mathring{g}_j,\mathring{v}_i,\mathring{f}_j), \tag{A2}$$

$$t^2: D(\mathring{g}_k,\mathring{v}_k,\mathring{v}_i) + \tfrac{1}{2}D(\mathring{g}_k,\mathring{f}_k,\mathring{g}_i) = D(\mathring{g}_j,\mathring{v}_i,\mathring{v}_j) + \tfrac{1}{2}D(\mathring{g}_j,\mathring{g}_i,\mathring{f}_j), \tag{A3}$$

$$t^3: D(\mathring{g}_k,\mathring{v}_k,\mathring{g}_i) = D(\mathring{g}_j,\mathring{g}_i,\mathring{v}_j) \tag{A4}$$

($i,j,k = 1,2,3$, i,j,k pairwise different; no summation over repeated indices here).

Equation (7b) yields

$$t^0: \epsilon_{ijk} D(\mathring{g}_i,\mathring{f}_j,\mathring{f}_k) = -8\pi G\rho_0 \Leftrightarrow \nabla_X \cdot \mathring{\mathbf{g}} = -4\pi G\rho_0, \tag{A5}$$

$$t^1: \epsilon_{ijk}(D(\mathring{g}_i,\mathring{f}_j,\mathring{v}_k) + D(\mathring{g}_i,\mathring{v}_j,\mathring{f}_k)) = 0, \tag{A6}$$

$$t^2: \epsilon_{ijk}(D(\mathring{g}_i,\mathring{f}_j,\mathring{g}_k) + D(\mathring{g}_i,\mathring{v}_j,\mathring{v}_k)) = 0, \tag{A7}$$

$$t^3: \epsilon_{ijk} D(\mathring{g}_i,\mathring{g}_j,\mathring{v}_k) = 0, \tag{A8}$$

$$t^4: D(\mathring{g}_1,\mathring{g}_2,\mathring{g}_3) = 0 \tag{A9}$$

($\epsilon_{ijk} = \epsilon_{[ijk]}$, $\epsilon_{123} = +1$, summation over repeated indices).

The advantage of the chosen form (7) of the equations is now obvious, since they yield compact functional expressions for the constraining equations.

Proof of the Lemma: Since each of Eqs. (A1)–(A4) yields three different equations we have 17 equations to solve in total. Starting with (A9) we get a functional relationship between $\mathring{g}_1,\mathring{g}_2,\mathring{g}_3$, which may be locally written as

$$\mathring{g}_3 = \psi_0(\mathring{g}_1,\mathring{g}_2), \tag{A10}$$

where ψ_0 is arbitrary.

Adding linear combinations of the five equations (A4)–(A8) we have

$$D(\mathring{g}_1,\mathring{g}_2,\mathring{v}_i) = 0, \quad i = 1,2,3, \tag{A11}$$

where (A10) has been used. Equation (A11) implies functional relations between $\mathring{g}_1$, $\mathring{g}_2$, and $\mathring{v}_i$, which are assumed to be solvable for v_i:

$$\mathring{v}_i = \psi_i(\mathring{g}_1,\mathring{g}_2), \quad i = 1,2,3. \tag{A12}$$

The ψ_i's are again arbitrary.

The assumption that (A12) is the solution to (A11) is only a weak restriction since it would not change the following procedure if the functional relations implied by (A11) were not solvable for $\mathring{v}_i$ at some points in the $(\mathring{g}_1,\mathring{g}_2,\mathring{v}_i)$ space, but for $\mathring{g}_1$ or $\mathring{g}_2$.

Linear combinations of (A3) added to (A7) yield, together with (A10)–(A12),

$$D(\mathring{g}_1,\mathring{g}_2,\mathring{f}_i) = 0, \quad i = 1,2,3. \tag{A13}$$

From Eq. (A13) we obtain for $i = 3$ with the help of (A1)

$$\mathring{g}_{1,1}\mathring{g}_{2,2} - \mathring{g}^2_{1,2} = 0, \tag{A14}$$

which implies locally the existence of a function A such that

$$\mathring{g}_2 = A(\mathring{g}_1). \tag{A15a}$$

Then (A10) and (A12) may be written as

$$\mathring{g}_3 = B(\mathring{g}_1), \tag{A15b}$$

$$\mathring{v}_1 = F(\mathring{g}_1), \tag{A15c}$$

$$\mathring{v}_2 = H(\mathring{g}_1), \tag{A15d}$$

$$\mathring{v}_3 = K(\mathring{g}_1), \tag{A15e}$$

where B, F, H, and K are arbitrary.

It is easy to convince oneself that the remaining equations [(A2) and (A6)] are trivially satisfied if (A15) holds.

Equations (A1) imply, by means of (A15a) and (A15b),

$$A'(\mathring{g}_1)\mathring{g}_{1,1} = \mathring{g}_{1,2}, \tag{A16a}$$

$$B'(\mathring{g}_1)\mathring{g}_{1,1} = \mathring{g}_{1,3}. \tag{A16b}$$

These differential equations are equivalent to

$$\frac{d\mathring{g}_1}{dX_2} = 0, \tag{A17a}$$

$$\frac{d\mathring{g}_1}{dX_3} = 0; \tag{A17b}$$

on curves,

$$\frac{dX_1}{dX_2} = -A', \tag{A18a}$$

$$\frac{dX_1}{dX_3} = -B'. \tag{A18b}$$

Since $\mathring{g}_1$ is constant on these curves, (A18a) and (A18b) integrate to

$$X_1 + A'X_2 + B'X_3 =: \xi, \quad \xi = \text{const}. \tag{A19}$$

Thus $\mathring{g}_1$ is an arbitrary function E of ξ:

$$\mathring{g}_1 = E(\xi). \tag{A20}$$

Q.E.D.

APPENDIX B: SOLUTION OF THE CONSTRAINTS (11a)–(11c)

Equations (12a) and (12b) are an extended version of the representation of solutions of $Z_{xx}Z_{yy} - Z^2_{xy} = 0$ for surfaces $Z(x,y)$ of vanishing Gaussian curvature given in Courant and Hilbert.[18]

We have to prove that the hypersurfaces $\mathring{\phi}$ given in (12a) and (12b) provide a representation of any solution of the set of equations (11a)–(11c). These equations imply the existence of functions A,B such that

$$\mathring{\phi}_{,2} = A(\mathring{\phi}_{,1}), \quad \mathring{\phi}_{,3} = B(\mathring{\phi}_{,1}) \tag{B1}$$

holds (see Appendix A).

Define a function C by ($\mathring{\phi}_{,1} =: \alpha$)

$$C := \mathring{\phi} - \mathring{\phi}_{,1}X_1 - \mathring{\phi}_{,2}X_2 - \mathring{\phi}_{,3}X_3 = \mathring{\phi} - \alpha X_1 - A(\alpha)X_2 - B(\alpha)X_3. \tag{B2}$$

Now we show that C is functionally dependent on α: The necessary and sufficient condition for two functions C and α of three variables X_1,X_2,X_3 to be functionally dependent is the vanishing of all subdeterminants of the matrix

$$\begin{bmatrix} C_{,1} & C_{,2} & C_{,3} \\ \alpha_{,1} & \alpha_{,2} & \alpha_{,3} \end{bmatrix}, \tag{B3}$$

i.e.,

$$C_{,2}\alpha_{,3} - C_{,3}\alpha_{,2} = 0,$$
$$C_{,1}\alpha_{,3} - C_{,3}\alpha_{,1} = 0, \qquad \text{(B4)}$$
$$C_{,1}\alpha_{,2} - C_{,2}\alpha_{,1} = 0.$$

Inserting (B2) into (B4) shows that these conditions are satisfied for any $\mathring{\phi}$ if (B1) holds.

Since C depends only on α we have

$$d\mathring{\phi} = \alpha\, dX_1 + A\, dX_2 + B\, dX_3 + (X_1 + A'X_2 + B'X_3 + C')d\alpha, \qquad \text{(B5)}$$

which implies

$$0 = X_1 + A'X_2 + B'X_3 + C'. \qquad \text{(B6)}$$

Here (B2) and (B6) are identical to the formulas (12a) and (12b); (B6) is also identical to (A20) with $-C'(\alpha) = E^{-1}(\alpha)$.

[1]A. Munier, Astrophys. J. **290**, 47 (1985), and references therein.
[2]E. N. Glass, J. Phys. A **13**, 3097 (1980), and references therein.
[3]R. N. Henriksen and M. De Robertis, Astrophys. J. **241**, 54 (1980).
[4]O. Heckmann and E. Schücking, (a) Z. Astrophys. **38**, 95 (1955); (b) Z. Astrophys. **40**, 81 (1956).
[5]J. D. Barrow and J. Silk, Astrophys. J. **250**, 432 (1981).
[6]J. A. Fillmore and P. Goldreich, Astrophys. J. **281**, 1 (1984).
[7]G. Lawitzky, Diplomarbeit, Max-Planck-Institut für Astrophysik, München, 1979.
[8]Ya. B. Zel'dovich, Astron. Astrophys. **5**, 84 (1970).
[9]S. F. Shandarin, A. G. Doroshkevich, and Ya. B. Zel'dovich, Usp. Fiz. Nauk. **139**, 83 (1983) [Sov. Usp. **26**, 46 (1983)].
[10]V. I. Arnol'd, S. F. Shandarin, and Ya. B. Zel'dovich, Geophys. Astrophys. Fluid Dyn. **20**, 111 (1982).
[11]M. Davies, G. Efstathiou, C. S. Frenk, and S. D. White, Astrophys. J. **292**, 371 (1985), and references therein.
[12]P. R. Shapiro and C. Struck-Marcell, Astrophys. J. Suppl. **57**, 205 (1985), and references therein.
[13]P. R. Woodward, Ann. Rev. Astron. Astrophys. **16**, 555 (1978), and references therein.
[14]R. Kates, Astron. Astrophys. **168**, 1 (1985).

[15]Here d/dt or equivalently an overdot denotes the total or Lagrangian derivative with respect to time and is defined by the operator $d/dt := \partial_t + \mathbf{v}\cdot\nabla$.
[16]The summation convention is adopted; a comma denotes a partial derivative with respect to the Lagrangian coordinates $i,j,k = 1,2,3$.
[17]This is except for the cylindrical surfaces $\mathring{\phi}(X_i = \text{const}, X_j, X_k)$ perpendicular to the (X_j, X_k) planes. Note that we have generalized the formula of Courant–Hilbert[18] for the three-dimensional case (see Appendix B).

[18]R. Courant and D. Hilbert, *Methoden der Mathematischen Physik* (Springer, Berlin, 1986), Bd. II, S. 8.

[19]Note that, in general, $\mathrm{II}(v_{i,k}) = \mathrm{II}(\theta_{jk}) + |\boldsymbol{\omega}|^2$ and $\mathrm{III}(v_{j,k}) = \mathrm{III}(\theta_{i,k})$, where the terms in parentheses indicate the tensor to which the invariants refer, which implies, in view of (19a), that the velocity gradient always has two vanishing eigenvalues.

[20]G. F. R. Ellis, "Relatisistic cosmology," in *Proceedings of Enrico-Fermi School*, Course XLVII, edited by B. K. Sachs (Academic, London, 1971).

[21]"Self consistency" is used in the sense of A. G. Doroshkevich, V. S. Ryaben'kii, and S. F. Shandarin, Astrofiz. **9**, 257 (1973) [Astrophys. **9**, 144 (1975)].

Nonlinear wave propagation in the Gauss–Bonnet extended Einstein theory

A. Tomimatsu
Department of Physics, Nagoya University, Nagoya 464, Japan
H. Ishihara
Department of Physics, Kyoto University, Kyoto 606, Japan

(Received 30 April 1987; accepted for publication 22 July 1987)

It is rather usual in general nonlinear fields that even if the waves begin to propagate under smooth initial conditions, after a finite time a shocklike discontinuity in the field quantities occurs at the wave front. Then the field theories fail to set a well-defined initial-value problem. By applying the method of characteristics, this propagation character of nonlinear gravitational waves which obey the lowest-order Gauss–Bonnet extended Einstein equations in higher-dimensional theory of gravity is studied. A diagonal metric tensor is assumed to be dependent on only two coordinate variables. It is found that the quadratic curvature terms added to the Einstein equations do not induce the occurrence of shock waves in the gravitational field.

I. INTRODUCTION

Although the correct theory is not yet known, most unified theories require more than four dimensions in space-time. The effective gravitational Lagrangian may contain higher-order curvature corrections in addition to the usual Einstein Lagrangian. Lovelock[1] proved that if the field equations are of second order the most general metric Lagrangian is simply expressed as a finite sum of dimensionally continued Gauss–Bonnet densities. Zwiebach[2] and Zumino[3] emphasized the importance of the Lovelock Lagrangian in terms of the ghost-free character of string theories. Even if the relevance to string theories is still obscure,[4] it is a natural Lagrangian to study higher-dimensional theories of gravity, and various properties of the cosmological or spherically symmetric solutions have been extensively investigated.[5] In this paper we propose to analyze another basic aspect of this extended gravity theory, i.e., nonlinear propagation of gravitational waves.

In the following we consider only the leading quadratic correction, and so the action has the form

$$I = I_1 + I_2, \tag{1}$$

$$I_1 = \int d^D x(-g)^{1/2} R/\kappa, \tag{2}$$

$$I_2 = \int d^D x(-g)^{1/2} \alpha (R_{\mu\nu\alpha\beta} R^{\mu\nu\alpha\beta} - 4R_{\mu\nu}R^{\mu\nu} + R^2), \tag{3}$$

in a D-dimensional ($D>4$) space-time with signature $(-,+,\ldots,+)$. We use conventions such that $R^{\mu}{}_{\nu\alpha\beta} = +\Gamma^{\mu}{}_{\nu\beta,\alpha}\cdots$, $R_{\mu\nu} = R^{\beta}{}_{\mu\beta\nu}$, $R = R^{\beta}{}_{\beta}$. The constant α is an undetermined parameter. One may be concerned with a small deviation $h_{\mu\nu}$ ($\equiv g_{\mu\nu} - \eta_{\mu\nu}$, $|h_{\mu\nu}| \ll 1$) from the flat metric $\eta_{\mu\nu}$. Then the quadratic Gauss–Bonnet action I_2 contributes only to terms quadratic and higher in $h_{\mu\nu}$ in the field equations, and no propagator correction is induced. If one considers nonlinear wave propagation, it is not so trivial to understand the contribution of I_2. What physical effects can be expected from the quadractic curvature terms?

Let us recall some results of the characteristic theory of the hyperbolic partial differential system in two independent variables with the form

$$U_{,t} + AU_{,x} + B = 0, \tag{4}$$

where U is a column vector with n components $u_1, u_2, \ldots, u_n$ and the matrix A and the column vector B are functions of field quantities u_i. We consider wave propagation in the positive direction of the x axis under smooth (Lipschitz continuous) initial conditions. It is rather usual in the nonlinear system that after a finite time t_c neighboring characteristic curves cross on the wave front and then a shocklike discontinuity in the field quantities occurs there.[6] We encounter an ill-posed Cauchy problem, because various solutions including the discontinuity (weak solutions) can satisfy the same initial conditions. In order to remove such nonuniqueness in the initial value problem, one must give a definite physical principle for selection of an adequate solution among many weak solutions.

We call a hyperbolic system "exceptional" if the critical time t_c does not exist. Semilinear equations in which A is a constant matrix become a typical example of exceptional systems. Our main concern here is to investigate whether in the lowest-order Gauss–Bonnet extended Einstein theory gravitational shock waves (discontinuities on wave fronts) occur under continuous initial conditions or not. It has been already known that the pure Einstein system in two independent variables is exceptional.[7] However the quadratic Gauss–Bonnet term makes nontrivial contributions to the matrix A in Eq. (4) and therefore the behavior of characteristic curves can be modified. The exceptional property in wave propagation may be lost.

In Sec. II we review briefly the method of characteristics for the general hyperbolic system (4) to derive a criterion for the occurrence of discontinuity.[6,7] With the aim of application of the criterion to the lowest-order Gauss–Bonnet extended Einstein system, we rewrite in Sec. III the field equations for gravitational wave propagation into the standard form (4). We assume a diagonal metric tensor dependent on

 0022-2488/87/112720-06$02.50

two coordinate variables t and x. Our conclusion obtained in Sec. IV is that in spite of the complicated form of matrix A this hyperbolic system including the quadratic curvature terms remains exceptional. The final section contains discussions of the problems left to future investigations.

II. CRITERION FOR OCCURRENCE OF DISCONTINUITY

For the hyperbolic system (4) all eigenvalues of matrix A are real. If the wave propagates in the positive direction of the x axis, there exists at least one positive eigenvalue. We denote one of the positive eigenvalues by $\lambda^{(\phi)}$, which is the velocity of wave propagation. The equation

$$\phi_{,t} + \lambda^{(\phi)}\phi_{,x} = 0 \tag{5}$$

determines the characteristic curve $\phi(x,t) = \text{const}$. If the multiplicity of the eigenvalue $\lambda^{(\phi)}$ is m_ϕ, there exist the corresponding left and right eigenvectors $l^{(\phi,k)}$ and $r^{(\phi,k)}$, where $k = 1,...,m_\phi$. The other left and right eigenvectors are denoted by $l^{(j)}$ and $r^{(j)}$, where $j = m_\phi + 1,...,n$.

Let us consider waves propagating into the constant state such that

$$B(U_0) = 0, \tag{6}$$

where U_0 is a constant solution of Eq. (4). The wave front is taken as

$$\phi(x,t) = 0, \tag{7}$$

which starts from $x = 0$ at $t = 0$ (see Fig. 1). Before a critical time U remains Lipschitz continuous on the wave front, i.e.,

$$\begin{aligned} &U(\phi = 0_-,t) = U(\phi = 0_+,t), \\ &\frac{\partial U(\phi = 0_-,t)}{\partial t} = \frac{\partial U(\phi = 0_+,t)}{\partial t}, \end{aligned} \tag{8}$$

where $\phi = 0_-$ and 0_+ mean limiting operations on the sides $\phi < 0$ (wave-disturbed region) and $\phi > 0$ (constant state), respectively. The derivatives with respect to ϕ admit a jump across the wave front as follows:

$$\frac{\partial U(\phi = 0_-,t)}{\partial \phi} - \frac{\partial U(\phi = 0_+,t)}{\partial \phi} \equiv \Pi(t) \neq 0,$$

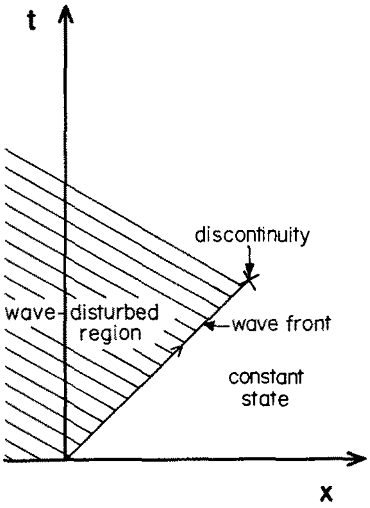

FIG. 1. Wave propagation into constant state.

$$\frac{\partial x(\phi = 0_-,t)}{\partial \phi} - \frac{\partial x(\phi = 0_+,t)}{\partial \phi} \equiv X(t) \neq 0, \tag{9}$$

where $\partial x(\phi,t)/\partial\phi$ ($\equiv x_{,\phi}$) represents a measure of distance between the two neighboring characteristic curves. If at $t = t_c$ the curve $\phi = \delta < 0$ in the disturbed side intersects with the wave front $\phi = 0$, we have

$$\frac{\partial x(\phi = 0_-,t_c)}{\partial \phi} = 0. \tag{10}$$

Then by virtue of the relation

$$U_{,x} = U_{,\phi}/x_{,\phi}, \tag{11}$$

the function U ceases to be Lipschitz continuous on the wave front.

Now we want to find the criterion for the occurrence of discontinuity. By differentiating the characteristic equation

$$\frac{\partial x(\phi,t)}{\partial t} = \lambda^{(\phi)} \tag{12}$$

with respect to ϕ, we obtain the result

$$X_{,t} = \sum_{i=1}^{n} \left(\frac{\partial \lambda^{(\phi)}}{\partial u_i}\right)_0 \Pi_i, \tag{13}$$

where Π_i is the ith component of the column vector Π. Hereafter the subscript 0 refers to the constant state solution. [In Eq. (13) $\partial\lambda^{(\phi)}/\partial u_i$ is assumed to be continuous on the wave front.] Integration of Eq. (13) gives

$$\frac{\partial x(\phi = 0_-,t_c)}{\partial \phi} = \frac{\partial x(\phi = 0_-,0)}{\partial \phi} + \sum_{i=1}^{n} \int_0^{t_c} \left(\frac{\partial \lambda^{(\phi)}}{\partial u_i}\right)_0 \Pi_i \, dt, \tag{14}$$

because the quantity $(x_{,\phi})_{\phi = 0_+}$ in Eq. (9) is constant. This tells us that Eq. (10) cannot be satisfied if

$$\sum_{i=1}^{n} \left(\frac{\partial \lambda^{(\phi)}}{\partial u_i}\right)_0 \Pi_i = 0. \tag{15}$$

The components Π_i are determined by multiplying Eq. (4) by the left eigenvectors $l^{(j)}$ and $l^{(\phi,k)}$. Using the jump conditions across the wave front, we have

$$\sum_{i=1}^{n} (l_0^{(j)})_i \Pi_i = 0, \quad j = m_\phi + 1,...,n, \tag{16}$$

$$\sum_{i=1}^{n} \left\{ (l_0^{(\phi,k)})_i \Pi_{i,t} + \left(\frac{\partial b^{(\phi,k)}}{\partial u_i}\right)_0 \Pi_i \right\} = 0, \quad k = 1,...,m_\phi, \tag{17}$$

where

$$b^{(\phi,k)} = \sum_{i=1}^{n} (l^{(\phi,k)})_i B_i.$$

From Eq. (16) we note that Π can be expressed as a linear combination of the right eigenvectors $r_0^{(\phi,k)}$. Thus Eq. (15) reduces to the final form of the criterion

$$\sum_{i=1}^{n} \left(\frac{\partial \lambda^{(\phi)}}{\partial u_i}\right)_0 (r_i^{(\phi,k)})_0 = 0, \tag{18}$$

which assures the exceptional property of the hyperbolic system. It is also true that if Eq. (18) does not hold the shock-like discontinuity can occur after a finite time. The criterion

(18) is very useful, since our remaining task is only to solve an eigenvalue problem of matrix A.

III. HYPERBOLIC EQUATIONS FOR GRAVITATIONAL WAVES

We now turn to the special hyperbolic system given by the quadratic Gauss–Bonnet extended Einstein theory. The nonlinear fields we consider here correspond to gravitational waves propagating in the direction of the x axis. We assume that the wave propagation is independent of the other spatial coordinates y^i $(i = 1,...,n)$. Then the line element in a $(n+2)$-dimensional space-time can be written in the form

$$ds^2 = -\exp f_0(x,t)(dt^2 - dx^2) + \sum_{i,j=1}^{n} h_{ij}(x,t)dy^i\,dy^j. \tag{19}$$

This metric may be interpreted as plane or cylindrical waves, but it differs from the so-called *pp* metric (plane-fronted parallel-ray waves) which has the form

$$ds^2 = 2H(u,y_i)du^2 + 2\,du\,dv + \sum_{i=1}^{n}(dy^i)^2. \tag{20}$$

The *pp* wave solution remains a solution of the pure Einstein equations, because then the Gauss–Bonnet contributions to the field equations vanish.[8] For the line element (19) the metric components h_{ij} rather then f_0 represent the dynamical degree of freedom of the gravitational field. The Gauss–Bonnet term survives even in Einstein geometries such that

$$G_{\mu\nu} \equiv R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R = 0. \tag{21}$$

In the following, for mathematical simplicity, the off-diagonal components of h_{ij} will be omitted. The metric contains only $n-1$ dynamical modes, since the trace part $\Sigma_{i=1}^{n}\ln h_{ii}$ obeys a constraint equation. We rewrite the diagonal components as

$$h_{ii} = \exp f_i. \tag{22}$$

The Einstein plus Gauss–Bonnet action yields field equations of second order. We express the equations as

$$G_{\mu\nu} + \kappa\alpha H_{\mu\nu} = 0, \tag{23}$$

where $G_{\mu\nu}$ is the Einstein tensor and $H_{\mu\nu}$ denotes the Gauss–Bonnet corrections. Each tensor can be divided into parts according to the degree in the second-order derivatives of f_a $(a = 0,1,...,n)$ as follows:

$$G_{\mu\nu} = G^{(0)}_{\mu\nu} + G^{(1)}_{\mu\nu}, \tag{24}$$

$$H_{\mu\nu} = H^{(0)}_{\mu\nu} + H^{(1)}_{\mu\nu} + H^{(2)}_{\mu\nu}. \tag{25}$$

The parts of the lowest degree, $G^{(0)}_{\mu\nu}$ and $H^{(0)}_{\mu\nu}$, are composed of f_a and the first-order derivatives only. We must obtain the matrix A in Eq. (4) to check the criterion (18). No explicit forms of these parts are necessary for this purpose.

The Einstein part linear in the second-order derivatives gives

$$G^{(1)}_{tt} = -\frac{1}{2}\sum_{i=1}^{n} f_{i,xx}, \tag{26}$$

$$G^{(1)}_{xx} = -\frac{1}{2}\sum_{i=1}^{n} f_{i,tt}, \tag{27}$$

$$G^{(1)}_{ii} = -\frac{1}{2}\sum_{a=0,\,a\neq i}^{n} (f_{a,tt} - f_{a,xx}). \tag{28}$$

Thus the pure Einstein equations become semilinear. This property is not altered even if the off-diagonal components of h_{ij} are present. It is obvious for the metric (19) that gravitational shock waves never occur under continuous initial conditions so far as the pure Einstein theory is concerned.

The Gauss–Bonnet linear part $H^{(1)}_{\mu\nu}$ includes very complicated functions of f_a, $f_{a,x}$, and $f_{a,t}$ in the coefficients of the second-order derivatives. A remarkable point for the Gauss–Bonnet extended Einstein equations is that there exist terms quadratic in the second-order derivatives such that

$$H^{(2)}_{tt} = H^{(2)}_{xx} = 0, \tag{29}$$

$$H^{(2)}_{ii} = \exp(-f_0)\sum_{\substack{j>k=1,\\ j\neq i,\ k\neq i}}^{n} (f_{j,tt}f_{k,xx} + f_{j,xx}f_{k,tt} - 2f_{j,tx}f_{k,tx}). \tag{30}$$

We regard the (x,x) and (i,i) components of Eqs. (23) as the set of equations which should be rewritten into the standard form (4). The solutions of this hyperbolic system automatically satisfy the remaining components of Eq. (23) if the initial conditions are adequately chosen.

The set of equations obtained here is not quasilinear, and so the application of the results in Sec. II seems impossible. However we can overcome this obstacle. Fortunately any terms quadratic in the second-order derivatives with respect to time, i.e., $f_{a,tt}f_{b,tt}$ are absent in $H^{(2)}_{ii}$. The set of equations can be considered as a linear inhomogenous algebraic system for $f_{a,tt}$. Then, by solving this algebraic system, it is always possible to express $f_{a,tt}$ as a function of the derivatives of lower order with respect to time, i.e.,

$$w_{a,t} = K_a(f_b, v_b, w_b, p_b, q_b), \tag{31}$$

where $v_b = f_{b,x}$, $w_b = f_{b,t}$, $p_b = f_{b,xx}$, and $q_b = f_{b,tx}$. Unless a discontinuity occurs, Eq. (31) admits a continuous extension of solutions from the initial state. Furthermore we find a manipulation which reduces Eq. (31) to a quasilinear system. We treat p_a and q_a as well as f_a, v_a, and w_a as independent unknown field variables, by giving the supplemental equations

$$f_{a,t} = w_a, \tag{32}$$

$$v_{a,t} = q_a, \tag{33}$$

$$p_{a,t} = q_{a,x}, \tag{34}$$

$$q_{a,t} = K_{a,x}, \tag{35}$$

besides Eq. (31). (If the function K_a is linear in p_a and q_a, the set of f_a, v_a, and w_a only should be treated as unknown field variables.) Now we arrive at the standard quasilinear form (4) in which the column vector U has $5(n+1)$ components $f_0,...,f_n$, $v_0,...,v_n$, $w_0,...,w_n$, $p_0,...,p_n$, and $q_0,...,q_n$, and matrix A is expressed as

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\delta_{ab} \\ 0 & -p_{ab} & -q_{ab} \end{pmatrix}. \tag{36}$$

Here we write the right-hand side of Eq. (35) as follows:

$$K_{a,x} = \sum_{b=0}^{n}\left[\left(\frac{\partial K_a}{\partial f_b}\right)v_b + \left(\frac{\partial K_a}{\partial v_b}\right)p_b + \left(\frac{\partial K_a}{\partial w_b}\right)q_b + p_{ab}\,p_{b,x} + q_{ab}q_{b,x}\right], \qquad (37)$$

$$p_{ab} = \frac{\partial K_a}{\partial p_b}, \quad q_{ab} = \frac{\partial K_a}{\partial q_b}. \qquad (38)$$

In the next section we will study the eigenvalue problem for matrix (36).

IV. ABSENCE OF GRAVITATIONAL SHOCK WAVES

Let us denote the A th component of right eigenvectors of matrix (36) by r_A $[A = 1,...,5(n+1)]$. The eigenvectors of the type

$$r_{3(n+1)+1} = \cdots = r_{5(n+1)} = 0$$

are uninteresting since the corresponding eigenvalue is equal to zero. The eigenvalue λ related to the dynamical modes is given by the equation

$$\det\begin{pmatrix} \lambda\delta_{ab} & \delta_{ab} \\ p_{ab} & q_{ab} + \lambda\delta_{ab} \end{pmatrix} = 0, \qquad (39)$$

and so it is sufficient to consider the eigenvectors of the type

$$r_1 = \cdots = r_{3(n+1)} = 0.$$

This means that the derivatives $\partial\lambda/\partial f_a$, $\partial\lambda/\partial v_a$, and $\partial\lambda/\partial w_a$ have nothing to do with the criterion (18).

Gravitational waves propagate into a constant state (6) that we assume to be the flat metic, where $f_a = v_a = w_a = p_a = q_a = 0$. The occurrence of dicontinuity is determined by the values of the derivatives $\partial\lambda/\partial p_a$ and $\partial\lambda/\partial q_a$ at the wave front where the unknown field variables $f_a,...,q_a$ are continuous. We can obtain these values by solving Eq. (39) under the conditions $f_a = v_a = w_a = 0$ which make the terms $H^{(1)}_{\mu\nu}, H^{(0)}_{\mu\nu}$, and $G^{(0)}_{\mu\nu}$ vanish. Our tedious task is to derive the explicit forms of the functions $K_a(f_b = 0, v_b = 0, w_b = 0, p_b, q_b)$ in Eq. (31). From the surviving terms of the (x,x) and (i,i) components in Eq. (23) we have the $n+1$ algebraic equations

$$\sum_{i=1}^{n} K_i = 0, \qquad (40)$$

$$\sum_{\substack{a=0,\\ a\neq i}}^{n} (K_a - p_a) - 2\kappa\alpha \sum_{\substack{j>k=1,\\ j\neq 1,\ k\neq i}}^{n} (K_j p_k + p_j K_k - 2q_j q_k) = 0. \qquad (41)$$

By subtracting the ith component of Eq. (41) from any other components, we can cancel the term $K_0 - p_0$ from this algebraic system. Hence the solution K_i becomes independent of both p_0 and q_0 (i.e., $p_{i0} = q_{i0} = 0$), and then the characteristic equation (39) reduces to

$$[\lambda(q_{00} + \lambda) - p_{00}] \times \det\begin{pmatrix} \lambda\delta_{ij} & \delta_{ij} \\ p_{ij} & q_{ij} + \lambda\delta_{ij} \end{pmatrix} = 0. \qquad (42)$$

Because we have $p_{00} = 1$ and $q_{00} = 0$, the first factor in the left-hand side of Eq. (42) gives the constant eigenvalues $\lambda = \pm 1$ which correspond to unphysical modes decoupled from any dynamical modes due to the metric components f_i. Thus the criterion (18) for our hyperbolic system should be read as

$$\sum_{\beta=1}^{2n}\left(\frac{\partial\lambda}{\partial u_\beta}\right)_0 (r_\beta)_0 = 0, \qquad (43)$$

where $u_1 = p_1,...,u_n = p_n$, $u_{n+1} = q_1,...,u_{2n} = q_n$, and r_β is the β th component of the right eigenvectors of the $2n\times 2n$ matrix

$$\begin{pmatrix} 0 & -\delta_{ij} \\ -p_{ij} & -q_{ij} \end{pmatrix}. \qquad (44)$$

The subscript 0 refers to the value at the wave front where $f_a = v_a = w_a = p_a = q_a = 0$.

Now let us obtain the eigenvalues of the matrix (44). The algebraic equations (40) and (41) admit the solution K_i of a power series in $\kappa\alpha$ as follows:

$$K_i = \sum_{s=0}^{\infty} (\kappa\alpha)^s K_i^{(s)}, \qquad (45)$$

where $K_i^{(s)}$ is a polynomial of order $s+1$ in p_i and q_i. Therefore the matrix (44) and its eigenvalue λ can be expressed in the same manner,

$$\begin{pmatrix} 0 & -\delta_{ij} \\ -\Sigma\,(\kappa\alpha)^s p_{ij}^{(s)} & -\Sigma\,(\kappa\alpha)^s q_{ij}^{(s)} \end{pmatrix}, \qquad (46)$$

$$\lambda = \sum_{s=0}^{\infty} (\kappa\alpha)^s \lambda_s. \qquad (47)$$

Since for the terms λ_s $(s\geqslant 2)$ of higher order in $\kappa\alpha$ all the derivatives $\partial\lambda_s/\partial u_\beta$ vanish at $p_i = q_i = 0$, we need only the first-order Gauss–Bonnet correction λ_1 if we want to evaluate the terms $(\partial\lambda/\partial u_\beta)_0$ in Eq. (43).

For the pure Einstein part, Eqs. (40) and (41) become

$$\sum_{i=1}^{n} K_i^{(0)} = 0, \qquad (48)$$

$$\sum_{a=0}^{n} (K_a^{(0)} - p_a) - (K_i^{(0)} - p_i) = 0. \qquad (49)$$

Equations (48) and (49) can be explicitly solved as follows:

$$K_i^{(0)} = p_i - \frac{1}{n}\sum_{j=1}^{n} p_j, \qquad (50)$$

which leads to

$$p_{ij}^{(0)} = \delta_{ij} - 1/n, \quad q_{ij}^{(0)} = 0. \qquad (51)$$

Thus we obtain the degenerate eigenvalues

$$\lambda_0 = 0, \pm 1.$$

The presence of the eigenvalue $\lambda_0 = 0$ results from the constraint equation for the trace part $\Sigma_{i=1}^{n} f_i$ [see Eq. (40)], while the eigenvalues $\lambda_0 = \pm 1$ (the multiplicity is $n-1$, respectively) represent gravitational waves propagating in the positive (or negative) direction of the x axis. Hence the components $(r_\beta)_0$ in Eq. (43) are given by the $n-1$ linearly independent right eigenvectors corresponding to $\lambda_0 = 1$ and satisfy the relations

$$(r_i)_0 = -(r_{i+n})_0 \equiv \gamma_i \quad (i = 1,2,...,n), \qquad (52)$$
$$\sum_{i=1}^{n} \gamma_i = 0.$$

The degeneracy of the unperturbed eigenvalue $\lambda_0 = 1$ is removed by the Gauss–Bonnet correction λ_1. The first-order terms in Eqs. (40) and (41) are

$$\sum_{i=0}^{n} K_i^{(1)} = 0, \qquad (53)$$

$$\sum_{\substack{a=0,\\ a\neq i}}^{n} K_a^{(1)} = 2\kappa\alpha \sum_{\substack{j>k=1,\\ j\neq i,\ k\neq i}}^{n} (K_j^{(0)} p_k + p_j K_k^{(0)} - 2q_j q_k)$$
$$= -2\sum_{j=1}^{n} p_j^2 + \frac{4}{n}\xi^2 - 2(1+2/n)\xi p_i + 4p_i^2 + 2\sum_{j=1}^{n} q_j^2 - 2\eta^2 + 4\eta q_i - 4q_i^2, \qquad (54)$$

where

$$\xi \equiv \sum_{i=1}^{n} p_i, \quad \eta \equiv \sum_{i=1}^{n} q_i. \qquad (55)$$

From Eqs. (53) and (54) we obtain

$$K_i^{(1)} = 4\Big\{ -p_i^2 + \frac{1}{n}\sum_{j=1}^{n} p_j^2 + \Big(\frac{1}{2}+\frac{1}{n}\Big)\Big(p_i - \frac{\xi}{n}\Big)\xi - \eta q_i + \frac{\eta^2}{n} + q_i^2 - \frac{1}{n}\sum_{j=1}^{n} q_j^2 \Big\}, \qquad (56)$$

then

$$p_{ij}^{(1)} = \frac{8}{n}p_j + \Big(2+\frac{4}{n}\Big)\Big(p_i - \frac{2}{n}\xi\Big) + \Big\{\Big(2+\frac{4}{n}\Big)\xi - 8p_i\Big\}\delta_{ij}, \qquad (57)$$
$$q_{ij}^{(1)} = -\frac{8}{n}q_j + \frac{8}{n}\eta - 4q_i + (8q_i - 4\eta)\delta_{ij}.$$

Following the usual perturbation method in degenerate cases, we evaluate the first-order correction λ_1 to the eigenvalue $\lambda_0 = 1$. This requires us to solve the algebraic equation

$$\det[X_{mm'} - \lambda_1 \delta_{mm'}] = 0, \qquad (58)$$

where $X_{mm'}$ is the $(n-1)\times(n-1)$ matrix defined by

$$X_{mm'} = l^{(m)} \begin{pmatrix} 0 & 0 \\ -p_{ij}^{(1)} & -q_{ij}^{(1)} \end{pmatrix} r^{(m')}$$
$$= \Big\{\Big(1+\frac{2}{n}\Big)\xi + 2\eta\Big\}\delta_{mm'} - 8\sum_{i=1}^{n}(p_i+q_i)\gamma_i^{(m)}\gamma_i^{(m')}, \qquad (59)$$

where $l^{(m)}$ and $r^{(m')}$ are $n-1$ left and right eigenvectors for $\lambda_0 = 1$ satisfying the orthonormal condition

$$\sum_{i=1}^{2n} l_i^{(m)} r_i^{(m')} = \delta_{mm'}.$$

In general, Eq. (58) will have $n-1$ roots which are expressed as

$$\lambda_1^{(k)} = f^{(k)}(\tau,\sigma_i) \quad (k = 1,2,...,n-1), \qquad (60)$$

where

$$\tau \equiv (1+2/n)\xi + 2\eta, \qquad (61)$$
$$\sigma_i \equiv p_i + q_i.$$

For example, in the case $n = 3$, the perturbed eigenvalue λ_1 turns out to be

$$\lambda_1 = \tau - \frac{4}{3}\sum_{i=1}^{3}\sigma_i \pm \frac{2}{3}\Big(\sum_{i\neq j=1}^{3}(\sigma_i - \sigma_j)^2\Big)^{1/2}. \qquad (62)$$

We derived the Gauss–Bonnet correction λ_1 to the velocity of wave propagation and found that it depends on the field variables p_i and q_i through the limited combinations of τ and σ_i. Now we are in a position to answer the main question in this paper: Does the field profile near the wave front steepen toward a shocklike discontinuity? The answer is no, because it is easy to check the orthogonality (43) with the aid of Eq. (52), i.e.,

$$\sum_{\beta=1}^{2n}\Big(\frac{\partial\lambda_1^{(k)}}{\partial u_\beta}\Big)_0 (r_\beta^{(m)})_0 = \sum_{i=1}^{n}\Big\{\frac{\partial f^{(k)}}{\partial p_i}\gamma_i^{(m)} - \frac{\partial f^{(k)}}{\partial q_i}\gamma_i^{(m)}\Big\}$$
$$= \sum_{i=1}^{n}\Big\{\Big(\frac{\partial f^{(k)}}{\partial\xi} + \frac{\partial f^{(k)}}{\partial\sigma_i}\Big) - \Big(\frac{\partial f^{(k)}}{\partial\eta} + \frac{\partial f^{(k)}}{\partial\sigma_i}\Big)\Big\}\gamma_i^{(m)}$$
$$= \Big(\frac{2}{n} - 1\Big)\frac{\partial f^{(k)}}{\partial\tau}\sum_{i=1}^{n}\gamma_i^{(m)} = 0. \qquad (63)$$

By virtue of this orthogonality relation the velocity λ of wave propagation satisfies the criterion (15). If we recall $U = 0$ at the side of the constant state (the flat metric), we get from Eq. (9),

$$\Pi = \frac{\partial U(\phi = 0_-,t)}{\partial\phi}, \qquad (64)$$

and so Eq. (15) gives

$$\frac{\partial\lambda(\phi = 0_-,t)}{\partial\phi} = \sum_i \Big(\frac{\partial\lambda}{\partial u_i}\Big)_0 \Pi_i = 0. \qquad (65)$$

Thus a characteristic curve at the side of wave-disturbed region $\phi = -|\delta|$ ($|\delta| \ll 1$) neighboring the wave front $\phi = 0$ becomes almost parallel to it, i.e.,

$$\lambda(\phi = -|\delta|) = \lambda(\phi = 0) + O(\delta^2). \qquad (66)$$

The intersection of the characteristic curve with the wave front is surely prohibited. The hyperbolic system in the lowest-order Gauss–Bonnet extended Einstein theory becomes exceptional so far as the diagonal metric for waves propagating into the flat space-time is concerned.

V. DISCUSSIONS

We have shown that the Gauss–Bonnet extended Einstein theory does not miss the property desirable for setting a well-defined initial value problem on wave propagation. However our proof in this paper covers only a limited case and many problems remain unsolved.

First, our proof should be extended to the case that the metric contains the off-diagonal components h_{ij} $(i\neq j)$. Even if it is improbable that the coupled terms, such as $h_{ij,xx} f_{i,tt}$, induce the occurrence of discontinuity, some definite proof will be necessary.

We assumed the flat metric in the undisturbed region. The theory has (anti-) de Sitter space as a solution. We can consider nonlinear wave propagation in (anti-) de Sitter space to check the occurrence of discontinuity there, although no straight application of the method mentioned in this paper is possible.

Let us give a comment on the contribution of the Gauss–Bonnet terms higher than quadratic in curvature. If all the terms do not contain the second-order derivatives with respect to time in nonlinear forms, such as $f_{i,tt} f_{j,tt}$, we can always rewrite the field equations into the standard quasilinear form (4) through the manipulation done in Sec. III. Then we obtain the eigenvalue λ of matrix A and express it as a power series in p_i and q_i. Because the Gauss–Bonnet combination of order k in curvature generates terms of at least order k in f_0 and f_i, only the quadratic curvature combination ($k = 2$) can contribute to the first-order correction λ_1 of the eigenvalue. This point is investigated in Ref. 9, and it is shown that even for the general Lovelock Lagrangian the above-mentioned conclusion is not altered.

If any quadratic and higher-order curvature terms different from the Gauss–Bonnet combinations are added to the gravitational Lagrangian, we encounter field equations of fourth order. The propagation character of gravitational waves is mainly determined by terms including the fourth-order derivatives (e.g., $R^{k-1}\Box R_{\mu\nu}$), and the Einstein Lagrangian plays no essential role in wave propagation. Apparently the field equations can be treated as a quasilinear system, and so the investigation of occurrence of shock waves is possible. A detailed comparison between the propagation characters of nonlinear waves in the higher-derivative theory of gravity and in the Gauss–Bonnet extended Einstein one is an interesting problem.

ACKNOWLEDGMENTS

The authors would like to thank Professor T. Taniuti for valuable discussions on nonlinear wave propagation.

This work was supported in part by the Grant in Aid for Cooperative Research (61302021) and for Encouragement of Scientist (61790094) from the Ministry of Education, Science, and Culture.

[1]D. Lovelock, J. Math. Phys. **12**, 498 (1971).
[2]B. Zwiebach, Phys. Lett. B **156**, 315 (1985).
[3]B. Zumino, Phys. Rep. **137**, 109 (1986).
[4]S. Deser and A. N. Redlich, Phys. Lett. B **176**, 350 (1986).
[5]See N. Deruelle and J. Madore, "Kaluza–Klein cosmology with the Lovelock Lagrangian," Institute Henri Poincaré preprint, 1986, and a list of references therein.
[6]A. Jeffrey and T. Taniuti, *Nonlinear Wave Propagation* (Academic, New York, 1964), p. 100.
[7]T. Taniuti, Prog. Theor. Phys. Suppl. **9**, 69 (1959).
[8]D. G. Boulware and S. Deser, Phys. Rev. Lett. **55**, 2656 (1985).
[9]A. Tomimatsu and H. Ishihara, Prog. Theor. Phys. **77**, 1014 (1987).

A case study of degeneracy in quantum statistics. I

William A. Barker, Daniel Raney, and John Sy
Department of Physics, Santa Clara University, Santa Clara, California 95053

(Received 15 April 1987; accepted for publication 22 July 1987)

In quantum statistics there are four categories of quantum degeneracy: nondegenerate, weak, intermediate, and strong. These are associated with the Fermi–Dirac and Bose–Einstein integrals, which are difficult to evaluate over the entire range of the activity parameter r defined as the particle density divided by the quantum concentration. In this paper (I), four classical systems with $r \ll 1$, and four weakly degenerate systems with $r \lesssim 1$ are examined. In the former, the Maxwell–Boltzmann distribution is sufficient. In the latter, the treatment of Landau and Lifschitz is extended. Physically realistic systems, like electrons in intrinsic and impurity semiconductors, and noble gases at different pressures and temperatures, are investigated. The neutrino and neutral pion systems are illustrative, albeit esoteric. Expressions for the quantum concentration for different dimensions and particle velocities are useful in predicting the onset of degeneracy. In the companion paper (II), intermediate and strong degeneracies are studied for $r \geq 1$.

I. INTRODUCTION

The word degenerate, in common usage, is pejorative. However, in physics, degenerate systems are among the most interesting and significant. In fact the word has two distinct scientific meanings.[1] A degenerate quantum system is one in which two or more wave functions share the same energy value. A degenerate statistical system is one in which a number of wave functions share the same region of space.

In the early days of modern physics, the word "degenerate" was used to describe abnormal or unusual behavior.[2] Thus it was considered normal for an eigenvalue to have just one associated eigenfunction. The Boltzmann distribution described a "normal" gas. The Fermi–Dirac and the Bose–Einstein distributions represented the degeneration of a gas.

In this paper (I) and the companion paper (II), we do a case study of degeneracy in statistical systems. Where possible, cases of real physical interest are selected. In some instances, for purposes of illustration, the physical systems described are fictitious.

II. THE SIGNIFICANT PARAMETERS IN DEGENERACY

In the literature, four related quantities are used to describe the degeneracy of a physical system; usually a gas of noninteracting particles. The absolute activity

$$r \equiv n/n_Q^*, \tag{1}$$

where n is the particle density and n_Q^* is the quantum concentration[3] n_Q times the quantum degeneracy g. (In these two papers we refer to n_Q^* as the quantum concentration.) If $r \ll 1$, Fermi and Bose gases behave classically. Otherwise the gas is in the quantum regime.

In systems where the concentration n is fixed and the temperature is variable, a degeneracy temperature T_0 is defined, using the condition that $r = 1$. Then

$$T \gg T_0 \quad \text{or} \quad T \lesssim T_0 \tag{2}$$

designates the classical or the quantum regime, respectively.

The chemical potential μ in Maxwell–Boltzmann (MB), Bose–Einstein (BE), and Fermi–Dirac (FD) statistics is a normalization parameter which depends on n and T. Although it is frequently less accessible than r or T_0, the quantity

$$\eta = \mu/kT \tag{3}$$

is often used to characterize the degree of degeneracy. In MB statistics η is always negative. In BE statistics η may be negative or zero. In FD statistics η may be negative, zero, or positive.

In the majority of calculations in statistical physics, an expression for the density of states $D(\epsilon)$ is required. This depends on the number of dimensions in the problem, and on the relativistic character of the particles. It is *independent* of the type of statistics used (MB, BE, or FD). A classical MB calculation using each of the density of states expressions enables one to find explicit formulas for the quantum concentration used in the ratio (1). We designate the density of states as the fourth parameter that is useful in describing all types of degeneracy.

We develop expressions for the density of states in one, two, or three dimensions for either nonrelativistic or ultrarelativistic particles,[4]

$$D_1(p)dp = 2gL/2\pi\hbar, \tag{4a}$$

$$D_2(p)dp = 2gA\pi p\, dp/(2\pi\hbar)^2, \tag{4b}$$

$$D_3(p)dp = gV4\pi p^2\, dp/(2\pi\hbar)^3. \tag{4c}$$

In (4), L is the length, A is the area, and V is the volume; if the particle rest mass $m \neq 0$, the quantum dengeneracy $g = 2S + 1$, where S is the spin; if $m = 0$, $g = 2S$, p is the momentum, and $\hbar$ is Planck's constant divided by 2π.

If the particles are nonrelativistic, $\epsilon = p^2/2m$ and the corresponding density of states expressions,[5] as a function of energy, are

$$D_1(\epsilon) = gL(2m/\epsilon)^{1/2}/2\pi\hbar, \tag{5a}$$

$$D_2(\epsilon) = gAm/2\pi\hbar^2, \tag{5b}$$

$$D_3(\epsilon) = gVm^{3/2}(2\epsilon)^{1/2}/2\pi^2\hbar^3. \tag{5c}$$

On the other hand, if the particles are ultrarelativistic, $\epsilon = cp$, in which case

$$D_1(\epsilon) = gL/\pi\hbar c, \quad (6a)$$

$$D_2(\epsilon) = gA\epsilon/2\pi\hbar^2c^2, \quad (6b)$$

$$D_3(\epsilon) = gV\epsilon^2/2\pi^2\hbar^3c^3, \quad (6c)$$

where c is the speed of light in a vacuum.

The quantum concentration n_Q^*, discussed in connection with (1), should be considered as a generic symbol. It could represent a linear, surface, or volume concentration. There are six different expressions for the quantum concentration, one for each of the density of states expressions. We calculate $n_Q^* = \lambda_Q^*$ from Eq. (5a). The other five functions may be found in a similar fashion,

$$N = \int_0^\infty D(\epsilon)e^{(\mu-\epsilon)/kt}\,d\epsilon$$

$$= \frac{e^{\mu/kT}gL}{(2\pi\hbar)(2m)^{1/2}}\int_0^\infty \epsilon^{-1/2}e^{-\epsilon/kt}\,d\epsilon. \quad (7)$$

Integrating and solving for $\eta = \mu/kT$, we find

$$\eta = \ln(\lambda/\lambda_Q^*), \quad (8)$$

where $\lambda = N/L$, and

$$\lambda_Q^* = g(mkT)^{1/2}/(2\pi\hbar^2)^{1/2}. \quad (9a)$$

Similarly, for nonrelativistic particles,

$$\sigma_Q^* = gmkT/2\pi\hbar^2, \quad (9b)$$

and

$$\rho_Q^* = g(mkT/2\pi\hbar^2)^{3/2}. \quad (9c)$$

For ultrarelativistic particles,

$$\lambda_Q^* = gkT/\pi\hbar c, \quad (10a)$$

$$\sigma_Q^* = (g/2\pi)(kT/\hbar c)^2, \quad (10b)$$

and

$$\rho_Q^* = (g/\pi^2)(kT/\hbar c)^3. \quad (10c)$$

The dimensions of λ_Q^*, σ_Q^*, and ρ_Q^* are cm^{-1}, cm^{-2}, and cm^{-3}, respectively. The density of states (5c) and (6c) and the corresponding quantum concentrations (9c) and (10c) are well known.[6] The other expressions may not be so familiar. All six of the quantities n_Q^* may be calculated from the appropriate Boltzmann partition function Z, where Z is of the form n_Q^*V'/g and V' represents a volume, area, or length.[7]

III. NONDEGENERATE SYSTEMS

A nondegenerate system is classical and is described for fermions and bosons alike by the Maxwell–Boltzmann distribution.

A. Fermions

Case 1: The conduction electrons in an intrinsic semiconductor,[8] e.g., Si.

In this three-dimensional problem, the conduction electron speed $v \ll c$, the effective mass m^* is 1.06 times the free electron mass, and $g = 2$. The energy of an electron $\epsilon - \epsilon_c$ refers to the conduction band energy ϵ_c as the origin. We use (5c) for the density of states. The particle density

$$\rho_e = \frac{2^{1/2}m^{*3/2}}{\pi^2\hbar^3}\exp\left[\frac{-(\epsilon_c-\mu)}{kT}\right]$$

$$\times\int_{\epsilon_c}^\infty (\epsilon-\epsilon_c)^{1/2}\exp\frac{-\epsilon+\epsilon_c}{kT}\,d\epsilon. \quad (11)$$

By integration, this becomes

$$\rho_e = 2\left(\frac{mkT}{2\pi\hbar^2}\right)^{3/2}\exp\left[\frac{\mu-\epsilon_c}{kT}\right]. \quad (12)$$

We define

$$\eta = (\mu-\epsilon_c)/kT = \ln(\rho_e/\rho_Q^*), \quad (13)$$

where

$$\rho_Q^* = 2(m^*kT/2\pi\hbar^2)^{3/2} = 3.87\times10^{19}\ \text{cm}^{-3},$$

$$\text{for } T = 300\ \text{K}. \quad (14)$$

If we wish to find η, we must consider the concentration of holes, as this quantity depends on the energy gap and the effective masses of electrons and holes,

$$\rho_h = 2(m_h^*kT/2\pi\hbar^2)^{3/2}\exp((\epsilon_v-\mu)/kT), \quad (15)$$

where $m_h^* = 0.58$ m in Si and ϵ_v is the energy at the top of the valence band. In an intrinsic semiconductor $\rho_e = \rho_h$. Using (12) and (15) we have

$$\eta = -\frac{\epsilon_g}{2kT} + \frac{3}{4}\ln\frac{m_h^*}{m_e^*}, \quad (16)$$

where $\epsilon_g = \epsilon_c - \epsilon_v$ is the energy gap. In Si, $\epsilon_g = 1.14$ eV. From (12), (13), (14), and (16) we find $\rho_e = 4.78\times10^9$ cm^{-3}, $r = \rho_e/\rho_Q^* = 1.24\times10^{-10}$, and $\eta = -22.8$. The degeneracy temperature follows when we set $r = 1$ and solve for T_0 from (14),

$$T_0 = 9.39\times10^{-5}\ \text{K}. \quad (17)$$

In case 1, we have seen that $r \ll 1$, η is negative and large, and $T_0 \ll T$. These values justify the use of MB statistics.

Case 2: A one-dimensional stream of neutrinos at $T = 300$ K, with a density of $\lambda = N_v/L = 5\times10^9\ \text{cm}^{-1}$.

The number density is numerically about the same as ρ_0 for the conduction electrons in Si. But of course, the dimensions differ. Neutrinos have a zero rest mass and are treated ultrarelativistically. For a neutrino $g = 1$. The appropriate density of states expression is (6a). The number of neutrinos

$$N_v = \frac{Le^{-\mu/kT}}{\pi\hbar c}\int_0^\infty e^{-\epsilon/kT}\,d\epsilon. \quad (18)$$

Integrating and solving for η yields

$$\eta = \ln(\lambda/\lambda_Q^*), \quad (19)$$

with λ_Q^* given by (10a). Numerically $\lambda_Q^* = 417\ \text{cm}^{-1}$. The absolute activity $r = 1.20\times10^7$. The calculation for λ_Q^* and r is correct since these quantities do not depend on the type of statistical distribution used. As $r \gg 1$, the appropriate statistical description is FD, not MB. The expression (19) for η is, therefore, invalid.

If we modify the problem dramatically and set $\lambda = 4\ \text{cm}^{-1}$ at $T = 300$ K, then $r = 9.57\times10^{-3}$ and $\eta = -4.65$. The degeneracy temperature

$$T_0 = (\pi\hbar c\lambda)/k = 2.87\ \text{K}. \quad (20)$$

This example is classical, but marginally so. The ratio r is about 1%, η is negative, but not large, and T is a factor of 104 above the degeneracy temperature.

From a physical point of view, a stream of neutrinos at $T = 300$ K, whose linear density $\lambda \sim 10\ \text{cm}^{-1}$ is much more likely than $\lambda \sim 10^9\ \text{cm}^{-1}$.

B. Bosons

Case 3: A three-dimensional gas of neon atoms at $T = 300$ K in a very high laboratory vacuum $p = 10^{-18}$ atm.[9] The mass of a Ne atom is 3.53×10^{-23} g.

We expect that the particles are nonrelativistic and that the statistical system is classical.

The particle speed may be found from

$$mv^2/2 = 3kT/2. \tag{21}$$

This gives a $v = 5.93 \times 10^4$ cm/sec which is indeed much smaller than c. The particle density follows from the equation of state

$$p = \rho kT. \tag{22}$$

The neon concentration

$$\rho = 24.5\ \text{cm}^{-3}. \tag{23}$$

Hence

$$r = \rho/\rho_Q^* = 2.55 \times 10^{-25}, \tag{24}$$

which is extremely small compared to unity. Therefore, the classical prescription may be used to find

$$\eta = \mu/kT = \ln r = -56.6. \tag{25}$$

As a final check in this case, we calculate the degeneracy temperature, which should be very small compared to 300 K,

$$T_0 = 2\pi\hbar^2\rho^{2/3}/mk = 1.21 \times 10^{-14}\ \text{K}. \tag{26}$$

Case 4: Neutral pions with a kinetic energy of 200 MeV and a volume density $\rho \leqslant 10^9\ \text{cm}^{-3}$.

The spin of the π^0 is zero. Hence $g = 1$. The rest mass[10] of this particle is $264.3 m_e = 135\ \text{MeV}/c^2$.

Since the kinetic energy exceeds the rest energy, the particles are relativistic. Indeed, using

$$m_0c^2(\gamma - 1) = \text{KE}, \tag{27}$$

with $\gamma = (1 - v^2/c^2)^{-1/2}$, it follows that $v = 0.92c$. We use the kinetic temperature

$$3kT/2 = \text{KE} \tag{28}$$

to calculate $T = 1.54 \times 10^{12}$ K. A comparison of ρ with ρ_Q^* will tell us whether or not the MB distribution is valid. From (10c),

$$\rho_Q^* = (kT/hc)^3\pi^2 = 3.12 \times 10^{37}\ \text{cm}^{-3}. \tag{29}$$

The absolute activity

$$r = \rho/\rho_Q^* \leqslant 3.21 \times 10^{-29} \tag{30}$$

and

$$\eta = \ln r \leqslant -65.6. \tag{31}$$

The degeneracy temperature

$$T_0 = \hbar c(\pi^2\rho)^{1/3}/k \leqslant 492\ \text{K}. \tag{32}$$

Each of the quantities given by (30)–(32) is consistent with the classical MB representation for an ultrarelativistic gas.

In each of the four cases considered, the most practical and significant quantity to calculate is the absolute activity r, defined by Eq. (1). This is the ratio of the actual density of the gas to the quantum concentration. If $r \ll 1$, the MB distribution may be used to find η. It is generally not possible to guess n_Q^* with any accuracy. This quantity depends on the dimensionality and relativistic nature of the problem. It is for this reason that Eqs. (9) and (10) are quite useful.

IV. WEAKLY DEGENERATE SYSTEMS

In most of the calculations in statistical mechanics, the systems are either nondegenerate or strongly degenerate. In this section and in paper II, we consider two regions which are in between: weak degeneracy and intermediate degeneracy. We establish a range for r in each of these four regions.

A weakly degenerate system is one in which degenerate behavior is just beginning.[11]

In the description of an ideal gas, whether it be one, two, or three dimensional, a very important relationship is the equation of state. Classically, this is

$$p'V' = NkT. \tag{33}$$

In this paper, the primed quantities on the left represent pressure p and volume V in three dimensions, force per unit length Q and area A in two dimensions, and force F and length L in one dimension.

A Fermi–Dirac or Bose–Einstein gas, which is weakly degenerate, is described by (33) multiplied by a series of decreasing correction terms. We generate these correction terms, following the procedure given by Landau and Lifschitz.[12] The method starts with the grand potential,[13]

$$\Omega = \mp kT \int_0^\infty \ln(1 \pm e^{(\mu - \epsilon)/kT}) D(\epsilon) d\epsilon. \tag{34}$$

The upper (lower) sign corresponds to FD (BE) statistics. We use the series

$$\ln(1 \pm w) = \pm w - \frac{w^2}{2} \pm \frac{w^3}{3} - \frac{w^4}{4} \pm \cdots, \tag{35}$$

to expand the integrand. We insert each of the six density of states expressions from Eqs. (5) and (6) and integrate over $x = \epsilon/kT$ term by term. The resulting expression for the grand potential has the form

$$\Omega = -NkT\left(1 \mp \frac{e^\eta}{2^s} + \frac{e^{2\eta}}{3^s} \mp \frac{e^{3\eta}}{4^s} + \cdots\right), \tag{36}$$

where $s = \frac{5}{2}$ for a nonrelativistic three-dimensional gas. The leading term, of course, is $\Omega_{\text{Boltzmann}}$. We use the classical formula $\exp(\eta) = r$, and write the result in terms of T and V. Now we make use of the fact that small increments in thermodynamic functions, expressed in terms of the corresponding variables, are equal,[14] e.g.,

$$(\delta u)_{t,\mu,V'} = (\delta F)_{T,V',N}. \tag{37}$$

Thus

$$F = F_0 + \delta F = F_0 + NkT\left(\pm \frac{r}{2^s} - \frac{r^2}{3^s} \pm \frac{r^3}{4^s} + \cdots\right). \tag{38}$$

We find p' from the thermodynamic relation $p' = -(\partial F/\partial V')_T$. Noting that

$$\frac{\partial r}{\partial V'} = -\frac{n}{V' n_Q^*} \quad (39)$$

the result is

$$p' = nkT\left(1 + \sum_{l=1}^{\infty} \frac{(-1)^{l+1} l r^l}{(l+1)^s}\right) \quad (40)$$

for FD statistics and

$$p' = nkT\left(1 - \sum_{l=1}^{\infty} \frac{l r^l}{(l+1)^s}\right) \quad (41)$$

for BE statistics.

The parameter s in these equations depends on the relativistic character of the problem and on the dimensionality. See Table I.

This procedure may seem circuitous since the grand potential in (33) is in fact PV. The equation of state, however, connects the variables P, V, and T. Therefore we must eliminate the parameter μ from (34) which we do by means of the classical approximation

$$N = \int_0^\infty D(\epsilon) d\epsilon / e[\epsilon - \mu/kT] \pm 1$$

$$\approx \int_0^\infty D(\epsilon) e^{-\epsilon/kT} e^{\mu/kT} d\epsilon \quad (42)$$

or $\mu = \mu_{\text{Boltzmann}}$.[15]

Consider the first correction term in (40) and (41) for the three-dimensional nonrelativistic case,

$$\delta P = \pm N^2 \pi^{3/2} \hbar^3 / 2gV^2 m^{3/2} (kT)^{1/2}. \quad (43)$$

This agrees with the published result.[16] Our results are more general.

We use these now to define the appropriate range for two of the four types of degeneracy. This definition is arbitrary, as there are five values of s, depending on geometry and particle speed.

If the gas is classical (or nondegenerate), the absolute activity

$$0 < r < 2^s \times 10^{-2}. \quad (44)$$

Hence for a three-dimensional NR gas, r must be less than 5.66×10^{-2}. If $r = 2^s \times 10^{-2}$, the equation of state differs by 1% from the classical value. In general, nondegeneracy requires r to be less than 0.16.

If the gas is weakly degenerate, the absolute activity

$$2^s \times 10^{-2} \leqslant r < 1. \quad (45)$$

The weakly degenerate expressions for the equation of state given by (40) and (41) are obtained by expansion of the integrand in (34). These three series converge if $\exp(\eta) = r < 1$.

TABLE I. Here s = power of $(l+1)$ in Eqs. (40) and (41).

	Nonrelativistic problem	Ultrarelativistic problem
One dimension	$\frac{3}{2}$	2
Two dimensions	2	3
Three dimensions	$\frac{5}{2}$	4

On the other hand, if the gas exhibits intermediate or strong degeneracy, $r \gg 1$. In Paper II, we establish a range for each of these two regions.

A. Fermions

Case 5: Phosphorus doped silicon at $T = 4.2$ K with a donor concentration of 9.18×10^{15} cm^{-3}.

As this is a three-dimensional nonrelativistic problem, the quantum concentration is given by

$$\rho_Q^* = g(mkT/2\pi\hbar^2)^{3/2} = 2.08 \times 10^{16} \text{ cm}^{-3}. \quad (46)$$

We chose the donor concentration to be 20% of this value. The concentration of conduction electrons is less than the donor concentration and may be calculated by means of the grand partition function,[17]

$$z = 1 + g \exp((\mu + |\epsilon_d|/kT)). \quad (47)$$

Here, $|\epsilon_d| = 45 \times 10^{-3}$ eV, is the energy required to ionize a phosphorus donor[18] in Si. Hence

$$\rho_e = f\rho_Q^*/(1 + 2r \exp(|\epsilon_d|/kT)), \quad (48)$$

using $g = 2$ and $\exp(\eta) = \rho_e/\rho_Q^* = r$. With $f = 20\%$ and $|\epsilon_d| \ll kT$, we find $r = 0.153$.

If we use the first two correction terms in the equation of state (40), with $s = \frac{5}{2}$, we find that the pressure exceeds the classical value by 2.40%. The contributions from the higher-order terms are negligible.

Here $\eta = \ln r = -1.88$, which is negative, but not as large as in cases 1–4.

One might expect that this case of weak degeneracy would have a degeneracy temperature not far removed from the sample temperature of 4.2 K. Indeed

$$T_0 = r^{2/3} T = 1.20 \text{ K}. \quad (49)$$

Case 6: A one-dimensional stream of neutrinos at $T = 1$ K with a density $\lambda = 1$ cm^{-1}.

This is an ultrarelativistic one-dimensional problem, similar to case 2. The quantum concentration

$$\lambda_Q^* = gkT/\pi\hbar c = 1.39. \quad (50)$$

The absolute activity

$$r = \lambda/\lambda_Q^* = 0.720, \quad (51)$$

$$\eta = -0.329, \quad (52)$$

and

$$T_0 = rT = 0.718 \text{ K}. \quad (53)$$

With r an appreciable fraction of unity, we have a case of weak degeneracy in which η is negative, but small, and T is only slightly greater than the degeneracy temperature.

The generalized pressure for this gas is a force. In (40), we use $s = 2$. We find that F exceeds its classical value by 9.93%. Since r is fairly close to 1, the series does not converge rapidly. We used the first eight terms.

B. Bosons

Case 7: Consider helium to be an "ideal" gas at $P = 1.2$ atm and $T = 1$ K.

The density $\rho = P/kT = 8.78 \times 10^{21}$ cm^{-3}. The quantum concentration may be found from (9c) with $g = 1$,

$\rho_Q^* = 1.50\times 10^{21}$ cm^{-3}. Therefore the absolute activity $r=\rho/\rho_Q^* = 5.84$, $\eta = \ln r = 1.76$, and $T_0 = r^{2/3}T = 3.25$ K.

Helium atoms are bosons requiring the use of Eq. (41) for the equation of state. Here the correction terms are all negative. They do not alternate in sign as in FD statistics. We find that the pressure is 14.4% less than the classical value. The eighth- and higher-order terms in (41) make negligible contributions.

Case 8: Consider case 4, with all the parameters the same, except $\rho = 9\times 10^{42}$ cm^{-3}, an unlikely value, chosen for illustrative purposes. This is a three-dimensional ultrarelativistic problem. The absolute activity $r = 0.289$, $\eta = -1.24$, and $T_0 = r^{1/2}T = 8.28\times 10^{11}$ K. These values characterize weak degeneracy. The pressure calculated from Eq. (41) is less than the classical value by only 1.83%. This example shows that degeneracy for an ultrarelativistic gas of pions requires a particle concentration in which the interparticle spacing is less than the range of the strong force. Clearly such a gas is not "ideal."

V. CONCLUSION

In quantum statistics, there are four degeneracy regimes: classical or nondegenerate, weak, intermediate, and strong. The reason why there is more than one category is primarily mathematical. There is no single analytic method which can be used to find solutions to the Fermi–Dirac and Bose–Einstein integrals over the entire range of the activity parameter r. In this and the succeeding paper, we extend the elegant methods of Landau and Lifschitz, Joyce and Dixon,[19] and Blankenbecler,[20] so that it is possible to obtain analytic solutions for $0<r<\infty$. In this paper, we examine eight cases: four are classical systems and four are weakly degenerate systems. The fermions studied are electrons in both intrinsic and impurity semiconductors and neutrinos of various linear particle densities. The bosons studied include Ne and He atoms with a particle density ranging from 20 cm^{-3} to 10^{21} cm^{-3}. High energy neutral pions, which are also bosons, are studied using widely disparate particle densities. The semiconductor and ideal gas systems are physically plausible. Certainly the high energy, high density, weakly degenerate π^0 system is physically unrealistic. This particular example makes an interesting point: The interparticle spacing must be comparable to the range of the nuclear force before the onset of degeneracy, for the pion system.

[1]See, for example, C. Kittle and H. Kroemer, *Thermal Physics* (Freeman, San Francisco, 1980), footnote p. 182.
[2]W. Pauli, *General Principles of Quantum Mechanics* (Springer, Berlin, 1980), p. 48; F. Hund, *The History of Quantum Theory* (Barnes and Noble, New York, 1974), p. 168; W. H. Cropper, *The Quantum Physics* (Oxford U. P., New York, 1970), p. 186; G. Troup, *Understanding Quantum Mechanics* (Methuen, London, 1968), p. 22. An excellent general reference of degeneracy in BE and FD statistics is given by D. A. McQuarrie, *Statistical Thermodynamics* (Harper and Row, New York, 1973), Chap. 10.
[3]See Ref. 1, p. 73.
[4]For background, see L. D. Landau and E. M. Lifschitz, *Statistical Physics* (Addison–Wesley, Reading, MA, 1958), p. 156.
[5]See, for example, Ref. 1, pp. 202 and 218.
[6]See W. A. Barker, J. Math. Phys. **27**, 302 (1986); Ref. 1, p. 218; Ref. 4, pp. 160 and 166.
[7]See Ref. 1, pp. 61 and 73.
[8]The parameters and general approach are taken from Ref. 1, pp. 355–363.
[9]D. Halliday and R. Resnick, *Fundamentals of Physics* (Wiley, New York, 1970), p. 428.
[10]R. C. Weast, *Handbook of Chemistry and Physics* (Chemical Rubber, Cleveland, 1977), F-270.
[11]See Ref. 4, p. 159.
[12]See Ref. 4, pp. 157–159.
[13]See also, A. Isihara, *Statistical Physics* (Academic, New York, 1971), p. 83.
[14]See Ref. 1, p. 70.
[15]In his treatment of the three-dimensional nonrelativistic case, Isihara uses (34) and (42) without using the classical expression for the chemical potential. See Ref. 13, p. 83.
[16]See Ref. 4, Eq. (55.15), p. 159; Ref. 13, Eq. 3.17, p. 83.
[17]See Ref. 1, p. 141.
[18]See Ref. 1, p. 368.
[19]W. B. Joyce and R. W. Dixon, Appl. Phys. Lett. **31**, 354 (1977).
[20]R. Blankenbecler, Am. J. Phys. **25**, 279 (1957).

Correlation functions in finite memory-time reservoir theory

Henk F. Arnoldus and Thomas F. George
Departments of Physics and Astronomy and of Chemistry, 239 Fronczak Hall, State University of New York at Buffalo, Buffalo, New York 14260

(Received 6 April 1987; accepted for publication 22 July 1987)

Interaction of a small system S with a large reservoir R amounts to thermal relaxation of the reduced system density operator $\rho_S(t)$. The presence of the reservoir enters the equation of motion for $\rho_S(t)$ through the reservoir correlation functions $f_{kl}(\tau)$ (defined in the text), which decay to zero for $\tau\to\infty$ on a time scale τ_c. Commonly, this τ_c is much smaller than the inverse relaxation constants for the time evolution of $\rho_S(t)$. Then a series of approximations can be made, which lead to a Markovian equation of motion for $\rho_S(t)$. In this paper the assumption of a small reservoir correlation time is removed. The equation of motion for $\rho_S(t)$ is solved, and it appears that the memory effect, due to $\tau_c\not\simeq 0$, can be incorporated in a frequency dependence of the relaxation operator $\widetilde{\Gamma}(\omega)$. Subsequently, (unequal-time) quantum correlation functions of two system operators are considered, where explicit expressions for (the Laplace transform of) the correlation functions are obtained. They involve again the relaxation operator $\widetilde{\Gamma}(\omega)$, which accounts for the time regression. Additionally it is found that an initial-correlation operator $\widetilde{\Upsilon}(\omega)$ arises, as a consequence of the fact that the equal-time correlation functions do not factorize as $\rho_S(t)$ times the reservoir density operator. It is pointed out that the frequency dependence of $\widetilde{\Gamma}(\omega)$ and the occurrence of a nonzero $\widetilde{\Upsilon}(\omega)$ both arise as a result of $\tau_c\not\simeq 0$, and should therefore be treated on an equal footing. Explicit evaluation of $\widetilde{\Gamma}(\omega)$ and $\widetilde{\Upsilon}(\omega)$ shows that their matrix elements can be expressed entirely in $\tilde{f}_{kl}(\omega)$, just as in the Markov approximation. Hence no essential complications appear if one should go beyond the limits of a small reservoir correlation time τ_c.

I. INTRODUCTION

In many practical cases the equation for the evolution of the density operator $\rho(t)$ of a quantum system assumes the general form

$$i\hbar\frac{d}{dt}\rho(t)=[H_S'+H_R+H_I',\rho(t)],\tag{1.1}$$

where H_S' and the H_R pertain to separated components S (= system) and R (= reservoir) of the entire configuration, and H_I' denotes an interaction between S and R. Probably the most familiar example is spontaneous decay of an excited atom in empty space. Then, H_S' equals the atomic Hamiltonian (internal structure), H_R represents the electromagnetic field, and H_I' is the dipole coupling between the atom and the electric component of the radiation field, which causes the spontaneous transitions. Since H_R has a large (infinite) number of eigenstates, an exact diagonalization of the complete Hamiltonian $H_S'+H_R+H_I'$ is intractable. The interest is, however, in the behavior of the atom, as it is determined by its interaction with the radiation field (vacuum or black-body radiation). Therefore, one introduces the reduced atomic (system) density operator by

$$\rho_S(t)=\mathrm{Tr}_R\,\rho(t),\tag{1.2}$$

where the trace runs over all states of the radiation field (the reservoir). The issue of reservoir, relaxation, or heat-bath theory is then to derive an accurate equation of motion for $\rho_S(t)$, in which the properties of R only enter as simple (and explicit) parameter functions. In the theory of spontaneous decay these are the Einstein coefficients and the Lamb shifts.

Most crucial for the development of a relaxation theory is the concept of a large reservoir. If the system S were not present, the reservoir would be in a (thermal equilibrium) state $\bar\rho_R$, which obeys

$$[H_R,\bar\rho_R]=0,\quad \bar\rho_R^\dagger=\bar\rho_R,\quad \mathrm{Tr}_R\,\bar\rho_R=1,\tag{1.3}$$

and it is assumed that the interaction between S and R does not substantially affect this reservoir state. Or more precisely, the state $\bar\rho_R$ changes a little due to the interaction with S, but the effect on the time evolution of the system density operator $\rho_S(t)$ is negligible. In the quoted example this implies that an atom in complete vacuum should decay in the same fashion as an atom in space with a single photon present. As a consequence of this large-reservoir assumption, we can factorize the density operator as

$$\rho(t)\simeq\rho_S(t)\bar\rho_R,\tag{1.4}$$

in places where the value of $\rho(t)$ determines the strength of the interaction.

In order to derive a relatively simple equation for $\rho_S(t)$, a sequence of additional approximations is usually made, which rely on the fact that the reservoir correlation time τ_c is short in comparison with the inverse relaxation constants $1/\Gamma$. The idea is as follows. One derives an equation for $\rho_S(t)$ which contains a quantity of the form $\langle R(t)R(0)\rangle$, with R a typical reservoir operator (for instance the electric field), and where the angular brackets indicate an average with the density operator $\bar\rho_R$, e.g.,

$$\langle\cdots\rangle=\mathrm{Tr}_R\,\bar\rho_R(\cdots)\,.\tag{1.5}$$

Due to the many eigenvalues $\hbar\omega$ of H_R and the large cutoff frequency ω_c, the function $\langle R(t)R(0)\rangle$ will decay to zero on a time scale of the order of $\tau_c \sim 1/\omega_c$. On the other hand, as a result of the interaction between S and R, the system density operator $\rho_S(t)$ will decay on a time scale $1/\Gamma$ (with Γ an Einstein coefficient, for instance), and in many cases the relation

$$\Gamma\tau_c \ll 1 \tag{1.6}$$

holds. The validity of (1.6) allows a series of approximations (see the Appendix), commonly referred to as the Markov approximation.

For spontaneous decay the restriction (1.6) is rigorously justified, and the equation of motion for $\rho_S(t)$ is known for more than a decade.[1-3] However, not every reservoir has a short correlation time. For instance, an adsorbed atom or molecule on a surface interacts with the substrate through phonon coupling (crystals) or electron-hole pair creation (metals). In the case of physisorbed atoms on a harmonic crystal, the Hamiltonian H'_S accounts for the kinetic and potential energy of the atom. The potential supports bound states, separated by $\sim 10^6$–10^8 MHz (infrared), which is resonant with the thermal excitations of the crystal (phonons). Mechanical coupling (vibrations) between the adsorbed atom and the lattice atoms gives rise to thermal relaxation of the adbond system.[4-6] Typical relaxation constants acquire an order of magnitude of 10^3–10^6 MHz, whereas the cutoff frequency (Debye frequency) is of the order of 10^6 MHz. For electron-hole pair formation the situation is even worse, where we have $\Gamma\tau_c \gtrsim 1$ so that a Markov approximation can never be justified.[7]

There exist many relaxation theories. Most notable are the projection techniques,[8-10] a Langevin formulation,[2] and, as we adopt here, a reservoir approach.[11-13] A feature of the quoted theories is that they all lead to the same result as soon as the Markov approximation is imposed. Several attempts have been undertaken to drop this Markov assumption.[14-16] To the best of our knowledge, however, a treatment free of inconsistencies and leading to explicit expressions (rather than formal expressions which cannot be evaluated) was never formulated.

II. RESERVOIR INTEGRAL

In this section we set up the notation and derive an integral of Eq. (1.1), which is appropriate for imposing the reservoir assumption. The first step is a redefinition of the system Hamiltonian. We recall that the interaction Hamiltonian H'_I is an operator in $S+R$ space, and therefore its reservoir average $\langle H'_I\rangle$ will be an operator in S space. In order to eliminate so-called secular terms, we define the new system and interaction Hamiltonians by

$$H_S = H'_S + \langle H'_I\rangle , \tag{2.1}$$

$$H_I = H'_I - \langle H'_I\rangle , \tag{2.2}$$

and the advantage of this rearrangement comes from the fact that the reservoir average of H_I equals zero. Explicitly,

$$\langle H_I\rangle = 0. \tag{2.3}$$

A compact and transparent representation of reservoir theory can be obtained with a Liouville-operator formalism. If we introduce the Liouvillians L_α by

$$L_\alpha\sigma = \hbar^{-1}[H_\alpha,\sigma], \quad \alpha = S,R,I, \tag{2.4}$$

which defines their action on an arbitrary operator σ in $S+R$ space, then the equation of motion (1.1) becomes

$$i\frac{d}{dt}\rho(t) = (L_S + L_R + L_I)\,\rho(t) . \tag{2.5}$$

For later purposes we mention a few properties of the Liouvillians. First, L_S and L_R commute, since they act on a different part of Liouville space. Second, L_R stands for a commutator, which implies the relation

$$\mathrm{Tr}_R\, L_R\sigma = 0, \tag{2.6}$$

for any σ. From $[H_R,\bar\rho_R] = 0$, Eq. (1.3), we find

$$e^{-iL_Rt}\bar\rho_R = \bar\rho_R, \tag{2.7}$$

and due to the shift of the interaction over its average, L_I obeys

$$\mathrm{Tr}_R\, L_I(\sigma_S\bar\rho_R) = 0. \tag{2.8}$$

Here and in the following, σ_S will indicate an arbitrary operator in S space.

An integral of Eq. (2.5) reads

$$\rho(t) = e^{-i(L_S+L_R)(t-t_0)}\rho(t_0) - i\int_{t_0}^{t} dt'\, e^{-i(L_S+L_R)(t-t')}L_I\rho(t') , \tag{2.9}$$

and substitution into Eq. (2.5) then yields

$$i\frac{d}{dt}\rho(t) = (L_S+L_R)\rho(t) + L_I e^{-i(L_S+L_R)(t-t_0)}\rho(t_0) - iL_I\int_{t_0}^{t} dt'\, e^{-i(L_S+L_R)(t-t')}L_I\rho(t') , \tag{2.10}$$

which is an exact integral of the equation of motion. If we subsequently take the trace over the reservoir states, the left-hand side becomes $i\,d\rho_S(t)/dt$, which equals the rate of change of the system density operator due to the free evolution [the term $L_S\rho_S(t)$ on the right-hand side] and the coupling to the reservoir (terms proportional to L_I). Hence the integral in Eq. (2.10) accounts for the relaxation of $\rho_S(t)$, and its value is proportional to the coupling strength. Therefore we can adopt the reservoir assumption, Eq. (1.4), on $\rho(t')$ in the integrand. We then find the equation of motion for $\rho_S(t)$ to be

$$i\frac{d}{dt}\rho_S(t) = L_S\rho_S(t) + \mathrm{Tr}_R\, L_I e^{-i(L_S+L_R)(t-t_0)}\rho(t_0) - i\,\mathrm{Tr}_R\, L_I\int_{t_0}^{t} dt'\, e^{-i(L_S+L_R)(t-t')} \times L_I(\rho_S(t')\bar\rho_R), \tag{2.11}$$

for $t \geqslant t_0$. It is important to note that the initial value $\rho(t_0)$ of the density operator (not the system part) remains present in the equation of motion for $\rho_S(t)$, in general. Equation (2.5) determines the time evolution of $\rho(t)$ for $t \geqslant t_0$, and the solution of Eq. (2.5) is fixed as soon as an initial value $\rho(t_0)$

is prescribed. Since $\rho(t_0)$ is not determined by the equation of motion, a further specification of the initial state $\rho(t_0)$ is necessary.

III. DENSITY OPERATOR

For finite memory-time reservoirs the choice of $\rho(t_0)$ is more than a matter of convenience. If the system has been in contact with the reservoir prior to t_0, then $\rho(t_0)$ is determined by its time evolution in the recent past $t<t_0$, and consequently the value of $\rho(t_0)$ is no longer arbitrary. As a solution, we simply define the instant of time t_0 as the time point at which the interaction L_I is switched on. We can then always take t_0 to be arbitrarily far into the past. For $t\leqslant t_0$ the reservoir is in its thermal-equilibrium state $\bar{\rho}_R$, and the system density operator $\rho_S(t)$ evolves independently of the reservoir. Therefore, we have for $t\leqslant t_0$

$$\rho(t)=\rho_S(t)\bar{\rho}_R. \tag{3.1}$$

Substitution into Eq. (2.11) and applying Eqs. (2.7) and (2.8) then shows that the term with $\rho(t_0)$ vanishes identically, due to the shift of the interaction Hamiltonian over its average. Then the equation of motion for $\rho_S(t)$ becomes

$$i\frac{d}{dt}\rho_S(t)=L_S\rho_S(t)-iTr_RL_I \times\int_{t_0}^{t}dt' e^{-i(L_S+L_R)(t-t')}L_I(\rho_S(t')\bar{\rho}_R), \tag{3.2}$$

for $t\geqslant t_0$.

Solving Eq. (3.2) is most easily done in the Laplace domain. If we define

$$\tilde{\rho}_S(\omega)=\int_{t_0}^{\infty}dt\, e^{i\omega(t-t_0)}\rho_S(t), \tag{3.3}$$

then the transformed equation of motion reads

$$(\omega-L_S)\tilde{\rho}_S(\omega)=i\rho_S(t_0)-i\,\mathrm{Tr}_R\,L_I \times[i/(\omega-L_S-L_R)] \times L_I(\tilde{\rho}_S(\omega)\bar{\rho}_R), \tag{3.4}$$

with solution

$$\tilde{\rho}_S(\omega)=[i/(\omega-L_S+i\tilde{\Gamma}(\omega))]\rho_S(t_0). \tag{3.5}$$

Here we introduced the relaxation operator $\tilde{\Gamma}(\omega)$ as

$$\tilde{\Gamma}(\omega)\sigma_S=\mathrm{Tr}_R\,L_I[i/(\omega-L_S-L_R)]L_I(\sigma_S\bar{\rho}_R), \tag{3.6}$$

which can equivalently be written as

$$\tilde{\Gamma}(\omega)\sigma_S=\mathrm{Tr}_R\,L_I\int_0^{\infty}d\tau\, e^{i(\omega-L_S-L_R)\tau}L_I(\sigma_S\bar{\rho}_R). \tag{3.7}$$

From Eq. (3.5) we see that $\tilde{\rho}_S(\omega)$, and thereby $\rho_S(t)$ for $t\geqslant t_0$, is determined by $\rho_S(t_0)$ only, and not by $\rho_S(t)$ for $t\leqslant t_0$. This is of course a result of assumption (3.1). The memory in the time evolution of $\rho_S(t)$ is displayed in the frequency dependence of $\tilde{\Gamma}(\omega)$. In the Appendix we show that $\tilde{\Gamma}(\omega)$ acquires a constant value (ω independent) in the Markov approximation.

From Eq. (3.7) we notice that $\tilde{\Gamma}(\omega)$ has the form of a Laplace transform

$$\tilde{\Gamma}(\omega)=\int_0^{\infty}d\tau\, e^{i\omega\tau}\Gamma(\tau), \tag{3.8}$$

where $\Gamma(\tau)$ is given by

$$\Gamma(\tau)\sigma_S=\mathrm{Tr}_R\,L_Ie^{-i(L_S+L_R)\tau}L_I(\sigma_S\bar{\rho}_R), \tag{3.9}$$

for $\tau\geqslant0$. Rewriting the equation of motion (3.2) in terms of $\Gamma(\tau)$ gives

$$i\frac{d}{dt}\rho_S(t)=L_S\rho_S(t)-i\int_{t_0}^{t}dt'\,\Gamma(t-t')\rho_S(t'), \tag{3.10}$$

which reveals that the time width of $\Gamma(\tau)$ (its decay time for $\tau\geqslant0$) equals the memory time of the reservoir-interaction term. It is the width of $\Gamma(\tau)$ which is usually termed the reservoir correlation time τ_c. Then it follows from Eq. (3.8) that the frequency width of $\tilde{\Gamma}(\omega)$ is of the order of $1/\tau_c$, and for $\tau_c\to0$, $\tilde{\Gamma}(\omega)$ becomes independent of ω.

The time evolution of $\rho_S(t)$ for $t\geqslant t_0$ will have little significance in general, which is partially due to the factorization at $t=t_0$. Due to the coupling to the reservoir, the density operator $\rho_S(t)$ will relax to a steady state (thermal equilibrium)

$$\bar{\rho}_S=\lim_{t\to\infty}\rho_S(t) \tag{3.11}$$

on a time scale $1/\Gamma$, as mentioned in the Introduction. Here, Γ denotes a typical matrix element of $\tilde{\Gamma}(\omega)$ [not of $\Gamma(\tau)$]. From the identity

$$\bar{\rho}_S=\lim_{\omega\to0}-i\omega\tilde{\rho}_S(\omega) \tag{3.12}$$

and Eq. (3.5), we find the equation for $\bar{\rho}_S$ to be

$$(L_S-i\tilde{\Gamma}(0))\bar{\rho}_S=0. \tag{3.13}$$

This shows that the long-time solution of $\rho_S(t)$ is determined by $\tilde{\Gamma}(\omega)$ at $\omega=0$. Furthermore, we notice that the dependence on the initial value $\rho_S(t_0)$ has disappeared in Eq. (3.13), which reflects that the memory of the preparation of the system at t_0 is erased.

IV. CORRELATION FUNCTION

Measurement of the steady-state density operator $\bar{\rho}_S$ of a physical system is tantamount to the determination of its relaxation constants, which are the matrix elements of $\tilde{\Gamma}(0)$, as displayed in Eq. (3.13). Dynamical properties of the system in contact with the reservoir, however, are reflected in the time evolution of $\rho_S(t)$ before it reaches its steady state $\bar{\rho}_S$. In view of Eq. (3.5), this transient behavior of $\rho_S(t)$ is governed by the frequency dependence of the relaxation operator $\tilde{\Gamma}(\omega)$. Besides the fact that a density operator is not amenable to direct observation, we also see from Eq. (3.5) that the details of $\rho_S(t)$ depend on the preparation of the system at $t=t_0$. Obviously, it is impossible to fix $\rho_S(t_0)$ (say, the wave function of an atom) at a single instant of time, and subsequently measure its evolution for $t>t_0$.

A standard method of obtaining dynamical information about a system is by observation of steady-state correlation functions of system operators, say X and Y. If we take arbitrarily t_0 as the instant of time at which the Schrödinger and Heisenberg pictures coincide, then the time dependence of X is given by

$$X(t) = e^{iL(t-t_0)}X, \tag{4.1}$$

where L indicates the Liouvillian $L_S + L_R + L_I$ of the entire system. Then $X(t_0) = X$ is an operator in S space only, but for $t > t_0$, $X(t)$ is an operator in $S + R$ space, due to L_I. Hence the time evolution of $X(t)$ carries information on the interaction with the reservoir. The correlation function of two operators X and Y is defined as the expectation value

$$\langle\langle X(t')Y(t)\rangle\rangle = \mathrm{Tr}\,\rho(t_0)X(t')Y(t) . \tag{4.2}$$

The double-bracket notation indicates an average with the full density operator of $S + R$, rather than with $\bar{\rho}_R$ only [Eq. (1.5)]. Transformation of Eq. (4.2) to the Schrödinger picture gives

$$\langle\langle X(t')Y(t)\rangle\rangle = \mathrm{Tr}\; Ye^{-iL(t-t')}(\rho(t')X), \tag{4.3}$$

or equivalently

$$\langle\langle X(t')Y(t)\rangle\rangle = \mathrm{Tr}\,Xe^{-iL(t'-t)}(Y\rho(t)) . \tag{4.4}$$

We notice that the initial time t_0 has disappeared in Eqs. (4.3) and (4.4), which already removes the ambiguities associated with the preparation of $\rho_S(t_0)$. A steady-state correlation function is now defined as $\langle\langle X(t')Y(t)\rangle\rangle$ with $t \gg t_0$, $t' \gg t_0$, and $t - t'$ fixed. Then the system is in state $\bar{\rho}_S$, which is time independent and a solution of Eq. (3.13). The time regression of the correlation functions (their $t - t'$ dependence) is governed by the same exponential that determines the time evolution of $\rho(t)$, and therefore we can extract dynamical properties of the system by an observation of the steady-state correlation functions.

Commonly, time regressions are not measured directly. For atoms or molecules on a solid substrate, for instance, one determines the spectral profile for the absorption of low-intensity monochromatic laser radiation with frequency ω. The spectral distribution as a function of ω and in the steady state is then given by expressions of the form

$$I(\omega) = \lim_{t\to\infty} \int_t^\infty dt'\, e^{i\omega(t'-t)}\langle\langle X(t')Y(t)\rangle\rangle, \tag{4.5}$$

which will further be referred to as the spectrum. It is the goal of this paper to evaluate $I(\omega)$ for a system in interaction with a finite memory-time reservoir.

V. SPECTRUM

From Eq. (4.5) we observe that we need the correlation function for $t' \geqslant t$, and therefore the representation (4.4) is most suitable. Then the occurring exponential is the same as for the time evolution of $\rho(t)$. If we introduce the Hilbert-space operator (Liouville-space vector)

$$A(t',t) = e^{-iL(t'-t)}(Y\rho(t)), \tag{5.1}$$

then the correlation function can be represented by

$$\langle\langle X(t')Y(t)\rangle\rangle = \mathrm{Tr}_S\, XA_S(t',t), \tag{5.2}$$

which only involves the system part $A_S(t',t) = \mathrm{Tr}_R\, A(t',t)$. In terms of the Laplace transform with respect to t',

$$\tilde{A}_S(\omega,t) = \int_t^\infty dt'\, e^{i\omega(t'-t)}A_S(t',t), \tag{5.3}$$

the spectrum attains the form

$$I(\omega) = \lim_{t\to\infty} \mathrm{Tr}_S\, X\tilde{A}_S(\omega,t) . \tag{5.4}$$

Differentiating Eq. (5.1) with respect to t' yields the equation of motion for $A(t',t)$,

$$i\frac{d}{dt'}A(t',t) = (L_S + L_R + L_I)A(t',t) , \tag{5.5}$$

which has to be solved for $t' \geqslant t$, with initial value

$$A(t,t) = Y\rho(t). \tag{5.6}$$

Equation (5.5) is identical to Eq. (2.5) for $\rho(t)$, and integrals can be found in the same way. The difference between a density operator and a correlation function is that for $\rho(t)$ we can choose the initial value $\rho(t_0)$ arbitrarily, whereas for $A(t',t)$ the initial value is unambiguously given by Eq. (5.6). This reflects the fact that $A(t',t)$ is essentially a two-time quantity. Its regression from t to t' is governed by Eq. (5.5) and its dependence on t enters through the initial condition, Eq. (5.6). The memory in the time regression, due to the finite reservoir correlation time, is of course the same as for the density operator and can be accounted for by the frequency-dependent relaxation operator $\tilde{\Gamma}(\omega)$, as we shall show below. As a second effect of a finite τ_c the density operator $\rho(t)$ in the initial value will carry a memory of its time evolution in the recent past. It is tempting to argue that we consider the steady state $t\to\infty$, so that the density operator $\rho(t)$ is constant in time. By the large-reservoir assumption we know that the reservoir remains in the state $\bar{\rho}_R$, whereas the system is in state $\bar{\rho}_S$ for $t\to\infty$. This would imply the replacement $\rho(t)\to\bar{\rho}_S\bar{\rho}_R$ in Eq. (5.6), which in turn would eliminate the explicit t dependence of $A(t',t)$, making the limit $t\to\infty$ in Eq. (5.4) trivial. We shall show that this procedure cannot be justified if τ_c is finite.

Since Eq. (5.5) is identical to Eq. (2.5) for $\rho_S(t)$, we can derive the appropriate integral along the same lines. The analog of Eq. (2.11) is

$$i\frac{d}{dt'}A_S(t',t) = L_SA_S(t',t) + \mathrm{Tr}_R\, L_Ie^{-i(L_S+L_R)(t'-t)}(Y\rho(t)) - i\,\mathrm{Tr}_R\, L_I \int_t^{t'} dt''\, e^{-i(L_S+L_R)(t'-t'')}L_I(A_S(t'',t)\bar{\rho}_R), \tag{5.7}$$

which contains $\rho(t)$ explicitly. Now we can substitute the right-hand side of Eq. (2.9) for $\rho(t)$ and take for $\rho(t_0)$ the value of $\rho_S(t_0)\bar{\rho}_R$. Then Eq. (5.7) becomes

$$i\frac{d}{dt'}A_S(t',t) = L_S A_S(t',t) + \mathrm{Tr}_R\, L_I e^{-i(L_S+L_R)(t'-t)} L_Y e^{-i(L_S+L_R)(t-t_0)}(\rho_S(t_0)\bar{\rho}_R)$$
$$- i\,\mathrm{Tr}_R\, L_I e^{-i(L_S+L_R)(t'-t)} L_Y \int_{t_0}^{t} dt''\, e^{-i(L_S+L_R)(t-t'')} L_I \rho(t'')$$
$$- i\,\mathrm{Tr}_R\, L_I \int_{t}^{t'} dt''\, e^{-i(L_S+L_R)(t'-t'')} L_I(A_S(t'',t)\bar{\rho}_R), \tag{5.8}$$

where we introduced the Liouvillian L_Y by

$$L_Y \sigma_S = Y\sigma_S, \tag{5.9}$$

in order to avoid notations with too many brackets. In the second term on the right-hand side of Eq. (5.8), the exponentials with L_R act only on $\bar{\rho}_R$, because L_R commutes with L_S and L_Y. Therefore they cancel, according to Eq. (2.7). The remaining two exponentials and L_Y affect only $\rho_S(t_0)$, and the result is some operator σ_S in S space. With Eq. (2.8) we then find that the whole term is identically zero. Considering the third term on the right-hand side, we notice that it has the form of a reservoir integral, as in Eq. (2.10), which implies that we can factorize $\rho(t'')$ here. Then we define a "density operator" $\rho_{SR}(t)$ of $S+R$ space by

$$\rho_{SR}(t) = \int_{t_0}^{t} dt'\, e^{-i(L_S+L_R)(t-t')} L_I(\rho_S(t')\bar{\rho}_R), \tag{5.10}$$

which allows us to write Eq. (5.8) as

$$i\frac{d}{dt'}A_S(t',t) = L_S A_S(t',t)$$
$$- i\,\mathrm{Tr}_R\, L_I e^{-i(L_S+L_R)(t'-t)} L_Y \rho_{SR}(t)$$
$$- i\,\mathrm{Tr}_R\, L_I \int_{t}^{t'} dt''\, e^{-i(L_S+L_R)(t'-t'')}$$
$$\times L_I(A_S(t'',t')\bar{\rho}_R). \tag{5.11}$$

Next we take the Laplace transform of Eq. (5.11), recalling that

$$A_S(t,t) = L_Y \rho_S(t), \tag{5.12}$$

as follows from Eq. (5.6), and rearrange the terms. We then obtain

$$\tilde{A}_S(\omega,t) = \frac{i}{\omega - L_S + i\tilde{\Gamma}(\omega)}\Big\{L_Y\rho_S(t)$$
$$- \mathrm{Tr}_R\, L_I \frac{i}{\omega - L_S - L_R} L_Y \rho_{SR}(t)\Big\}. \tag{5.13}$$

The factor in front of the curly brackets is the same as in Eq. (3.5), and it represents the time regression from t to t' of $A_S(t',t)$. The first term inside the curly brackets, $L_Y\rho_S(t)$, corresponds to a factorized initial state. If we would have replaced $A(t,t) = Y\rho(t)$ by $Y(\rho_S(t)\bar{\rho}_R)$, then it is easy to see that the second term on the right-hand side of Eq. (5.7) would have disappeared, and thereby the second term in curly brackets in Eq. (5.13). Conversely, the term with $\rho_{SR}(t)$ in Eq. (5.13) accounts for the correlations between S and R in $\rho(t)$, which are present at the initial time for the time evolution of $A(t',t)$ from t to t'.

The explicit time dependence of $\tilde{A}_S(\omega,t)$ enters through $\rho_S(t)$ and $\rho_{SR}(t)$. If we denote their steady-state values by an overbar, then the spectrum, Eq. (5.4), becomes

$$I(\omega) = \mathrm{Tr}_S\, L_X \frac{i}{\omega - L_S + i\tilde{\Gamma}(\omega)}$$
$$\times\left\{L_Y\bar{\rho}_S - \mathrm{Tr}_R\, L_I \frac{i}{\omega - L_S - L_R} L_Y \bar{\rho}_{SR}\right\}, \tag{5.14}$$

with

$$L_X\sigma_S = X\sigma_S. \tag{5.15}$$

Expression (5.14) involves the system-reservoir state $\bar{\rho}_{SR}$, which might seem cumbersome. From Eq. (5.10) we find the Laplace transform of $\rho_{SR}(t)$ to be

$$\tilde{\rho}_{SR}(\omega) = [i/(\omega - L_S - L_R)]L_I(\tilde{\rho}_S(\omega)\bar{\rho}_R), \tag{5.16}$$

in terms of $\tilde{\rho}_S\omega$ from Eq. (3.5). Then the steady-stae $\bar{\rho}_{SR}$ follows from the identity (3.12), which gives

$$\bar{\rho}_{SR} = [i/(i0^+ - L_S - L_R)]L_I(\bar{\rho}_S\bar{\rho}_R). \tag{5.17}$$

Here, the notation $i0^+$ indicates a small positive imaginary part, which is necessary to assure the convergence of Laplace-transform integrals, or equivalently, the existence of the inverse of $i0^+ - L_S - L_R$. In the next section we show how to evaluate the right-hand side of Eq. (5.17). If we define an operator $\tilde{\Upsilon}(\omega)$ by

$$\tilde{\Upsilon}(\omega)\sigma_S = \mathrm{Tr}_R\, L_I \frac{i}{\omega - L_S - L_R}$$
$$\times L_Y \frac{1}{i0^+ - L_S - L_R} L_I(\sigma_S\bar{\rho}_R), \tag{5.18}$$

then the spectrum attains the form

$$I(\omega) = \mathrm{Tr}_S\, L_X[i/(\omega - L_S + i\tilde{\Gamma}(\omega))](L_Y - i\tilde{\Upsilon}(\omega))\bar{\rho}_S. \tag{5.19}$$

Equation (5.19) is the most condensed and general representation of the result of this paper. The finite memory time of the reservoir appears as a frequency dependence of the relaxation operator $\tilde{\Gamma}(\omega)$, and as a nonvanishing initial-correlation operator $\tilde{\Upsilon}(\omega)$.

VI. $\mathcal{S}\,\mathcal{R}$ INTERACTION

Although the result (5.19) is appealing and explicit, the occurring operators $\tilde{\Gamma}(\omega)$ and $\tilde{\Upsilon}(\omega)$, which represent the interaction of the system with reservoir, might look awkward in their definitions, Eqs. (3.6) and (5.18). Especially the reservoir Liouvillian L_R in denominators and the appearance of $i0^+$ in Eq. (5.18) might seem to make an explicit evaluation of $\tilde{\Gamma}(\omega)$ and $\tilde{\Upsilon}(\omega)$ intractable. Such is, however, not the case, as we shall show in this section.

Obviously, an elaboration of $\tilde{\Gamma}(\omega)$ and $\tilde{\Upsilon}(\omega)$ requires additional specifications of the interaction Hamiltonian H_I. It will turn out to be sufficient to assume the form

$$H_I = \hbar \sum_k \mathscr{S}_k \mathscr{R}_k, \tag{6.1}$$

with $\mathscr{S}_k$ ($\mathscr{R}_k$) a pure S (R) operator. The form (6.1) pertains to most practical situations we have encountered. In the case of fluorescence, $\mathscr{S}_k$ signifies the k th Cartesian component of the atomic dipole moment, and for adsorbates on a substrate the subscript k takes on two values, corresponding to the two terms in the binding (Morse) potential. In fact, the form (6.1) for H_I can always be enforced by an expansion in matrix elements.

Evaluation of the relaxation operator $\tilde{\Gamma}(\omega)$ starts from its representation (3.9) in the time domain. We expand the two L_I's as commutators, which gives rise to four terms. Then we insert H_I from Eq. (6.1), and we notice that every factor is an operator in S or R space only. Combining the R operators and taking the trace over the reservoir states then shows that the R contribution can be accounted for by a single complex-valued function

$$f_{kl}(\tau) = \langle \mathscr{R}_k e^{-iL_R\tau} \mathscr{R}_l \rangle, \tag{6.2}$$

which will be called the reservoir correlation function. We find

$$\Gamma(\tau)\sigma_S = \sum_{kl} L_l e^{-iL_S\tau} (f_{lk}(\tau)\mathscr{S}_k \sigma_S - f^*_{lk}(\tau)\sigma_S \mathscr{S}_k), \tag{6.3}$$

with

$$L_l \sigma_S = [\mathscr{S}_l, \sigma_S]. \tag{6.4}$$

Expression (6.3) only involves the system operators $\mathscr{S}_k$ and the Liouvillian L_S for the free evolution of the system. The reservoir enters via the parameter functions $f_{kl}(\tau)$, which can be found as soon as a particular reservoir is prescribed. For a harmonic crystal, for instance, the reservoir correlation functions are given analytically in Ref. 17.

The initial correlation operator $\tilde{\Upsilon}(\omega)$ from Eq. (5.18) is the Laplace transform of

$$\Upsilon(\tau)\sigma_S = \mathrm{Tr}_R\, L_I e^{-i(L_S + L_R)\tau} \times L_Y [1/(i0^+ - L_S - L_R)] L_I(\sigma_S \bar{\rho}_R). \tag{6.5}$$

First we recall that the notation $i0^+$ should be read as

$$\frac{1}{i0^+ - L_S - L_R} L_I(\sigma_S \bar{\rho}_R) = -i \lim_{\omega \to i0^+} \int_0^\infty d\tau' e^{i(\omega - L_S - L_R)\tau'} L_I(\sigma_S \bar{\rho}_R). \tag{6.6}$$

Then we insert the form (6.1) for the interaction and rearrange the S and R terms. We then obtain for $\Upsilon(\tau)$

$$\Upsilon(\tau)\sigma_S = i \sum_{kl} L_l e^{-iL_S\tau} L_Y \int_o^\infty dt' e^{-iL_S\tau'} \times (f_{lk}(\tau + \tau')\mathscr{S}_k \sigma_S - f^*_{lk}(\tau + \tau')\sigma_S \mathscr{S}_k), \tag{6.7}$$

where the reservoir is again entirely incorporated in the functions $f_{kl}(\tau)$. Since the $f_{lk}(\tau)$'s decay to zero sufficiently fast for $\tau \to \infty$, we omitted at this stage the $i0^+$ in the rightmost exponential.

VII. LAPLACE TRANSFORM

Before we can take the Laplace transform of Eqs. (6.3) and (6.7), we must work out the exponentials. Eigenstates of the system Hamiltonian will be denoted by $|a,\alpha\rangle$, where a indicates the energy and α any degeneracy. By definition they obey

$$H_S|a,\alpha\rangle = \hbar\omega_a |a,\alpha\rangle. \tag{7.1}$$

With respect to its own eigenstates we can write H_S as

$$H_S = \sum_{a,\alpha} \hbar\omega_a |a,\alpha\rangle\langle a,\alpha| = \sum_a \hbar\omega_a P_a, \tag{7.2}$$

with

$$P_a = \sum_\alpha |a,\alpha\rangle\langle a,\alpha|, \tag{7.3}$$

the projector on the subspace with energy $\hbar\omega_a$. From the orthonormality of the states $|a,\alpha\rangle$ we have

$$P_a P_b = \delta_{ab} P_a, \tag{7.4}$$

and from the completeness of the set $|a,\alpha\rangle$ we find the closure relation

$$\sum_a P_a = 1. \tag{7.5}$$

Then it is an easy matter to expand the exponential $\exp(-iL_S\tau)$ in projectors, which gives

$$e^{-iL_S\tau}\sigma_S = \sum_{ab} e^{-i\Delta_{ab}\tau} P_a \sigma_S P_b, \tag{7.6}$$

in terms of the level separations

$$\Delta_{ab} = \omega_a - \omega_b. \tag{7.7}$$

Next we substitute Eq. (7.6) into Eq. (6.3) and evaluate the Laplace transform. We obtain

$$\tilde{\Gamma}(\omega)\sigma_S = \sum_{kl} L_l \sum_{ab} P_a (\tilde{f}_{lk}(\Delta_{ba} + \omega)\mathscr{S}_k \sigma_S - \tilde{f}^*_{lk}(-\Delta_{ba} - \omega)\sigma_S \mathscr{S}_k) P_b, \tag{7.8}$$

in terms of the Laplace transform $\tilde{f}_{kl}(\omega)$ of $f_{kl}(\tau)$. Because L_l equals the commutator with $\mathscr{S}_l$, the right-hand side only involves operators $\mathscr{S}_l$ and projectors. If we insert the closure relation (7.5) in various places in Eq. (7.8), we immediately find the matrix representation of $\tilde{\Gamma}(\omega)$ in terms of matrix elements of $\mathscr{S}_l$. The result (7.8) is the most compact representation of the explicit form of $\tilde{\Gamma}(\omega)$.

In the very same way we find the Laplace transform of $\Upsilon(\tau)$ from Eq. (6.7), although with considerably more effort, which is due to the double integral (over τ and τ'). The result is

$$\tilde{\Upsilon}(\omega)\sigma_S = \sum_{kl} L_l \sum_{abc} \frac{1}{\Delta_{ac}+\omega} P_c Y P_a \times\{(\tilde{f}_{lk}(\Delta_{ba}) - \tilde{f}_{lk}(\Delta_{bc}+\omega))\mathscr{S}_k\sigma_S - (\mathscr{S}\tilde{f}^*_{lk}(-\Delta_{ba}) - \mathscr{S}\tilde{f}^*_{lk}(-\Delta_{bc}-\omega))\sigma_S\mathscr{S}_k\}P_b, \quad (7.9)$$

which has a striking resemblance with Eq. (7.8). Most remarkable is that $\tilde{\Upsilon}(\omega)$ can again be expressed in the reservoir correlation function $\tilde{f}_{kl}(\omega)$ which also determines $\tilde{\Gamma}(\omega)$, and, as shown in the Appendix, the relaxation operator in the Markov approximation. The distinction is that the functions $\tilde{f}_{kl}(\omega)$ occur with different arguments.

If ω equals a level separation $\Delta_{ca} = -\Delta_{ac}$, then the denominator of the first factor under the triple summation in Eq. (7.9) becomes zero. For $\omega = \Delta_{ca}$ we have $\Delta_{bc}+\omega = \Delta_{ba}$, and hence the difference of the two functions $\tilde{f}_{lk}$ in curly brackets also approaches zero. In the process of deriving Eq. (7.9) we found that this feature does not constitute a problem. The limit is simply

$$\lim_{\omega\to-\Delta_{ac}} \frac{1}{\Delta_{ac}+\omega}(\tilde{f}_{lk}(\Delta_{ba}) - \tilde{f}_{lk}(\Delta_{bc}+\omega)) = -\frac{d}{d\omega}\tilde{f}_{lk}(\omega), \quad \text{in } \omega = \Delta_{ba}, \quad (7.10)$$

and there is no singularity or discontinuity if ω passes across a resonance.

VIII. CONCLUSIONS

If the decay time τ_c of the reservoir correlation function $f_{kl}(\tau)$ for $\tau\to\infty$ is not small in comparison with the relaxation times $1/\Gamma$, which are determined by the same function [see Eq. (7.8)], then a Markov approximation cannot be correct. In this paper we imposed no limits on τ_c. We only assumed that the system S is small in comparison with the reservoir R. The finite value of τ_c amounts to a memory in the time evolution of the density operator, which is reflected in a frequency dependence of the relaxation operator $\tilde{\Gamma}(\omega)$. Correlation functions of system operators depend on two times, t' and t. The regression from t to t' exhibits the same memory effect as the time evolution of the density operator. Additionally, the equal-time correlation function, which is the initial value for the time regression, carries a memory to the recent past. It appears that this second phenomenon could be accounted for by an initial correlation operator $\tilde{\Upsilon}(\omega)$ in the expression for the spectrum $I(\omega)$.

Frequency-dependent relaxation operators are widely applied in the literature. Their Laplace inverse $\Gamma(\tau)$ is sometimes called a memory kernel, because it is the finite time width of $\Gamma(\tau)$ which brings about the memory in the time evolution, as is most obvious from Eq. (3.10). Initial-correlation operators, however, are rare.[18] Despite the fact that the frequency dependence of $\tilde{\Gamma}(\omega)$ originates from the same memory mechanism which amounts to a nonvanishing $\tilde{\Upsilon}(\omega)$, the latter is usually not found. As pointed out in the derivation of $\tilde{\Upsilon}(\omega)$, the disappearance of $\tilde{\Upsilon}(\omega)$ is a consequence of a factorization of the initial value or state, which cannot be justified in general.

ACKNOWLEDGMENTS

This research was supported by the Air Force Office of Scientific Research (AFSC), United States Air Force, under Contract F49620-86-C-0009, and the Office of Naval Research.

APPENDIX: MARKOV APPROXIMATION

It is illuminating to compare the results of the present paper with their equivalents under the Markov approximation. Since we already have Eq. (2.11) we can start here. First, we state that we are not interested in a time evolution of $\rho_S(t)$ on a time scale τ_c, which implies that we can factorize $\rho(t_0)$ as $\rho_S(t_0)\bar{\rho}_R$. In this fashion we discard the memory of the initial state to its past, which immediately gives $\tilde{\Upsilon}(\omega)\equiv 0$, or equivalently, the second term on the right-hand side of Eq. (2.11) is zero. Second, we know that if we work out the integral in Eq. (2.11), we find reservoir correlation functions $f_{kl}(t-t')$, which decay to zero on a time scale τ_c. Therefore, the major contribution to the integral comes from $t-\tau_c \lesssim t' \leqslant t$. Because we impose the condition $\Gamma\tau_c \ll 1$, the density operator $\rho_S(t')$ in the integrand is not affected significantly by the relaxation process on this small time interval. Then we can replace $\rho_S(t')$ by its free evolution

$$\rho_S(t') = e^{-iL_S(t'-t)}\rho_S(t), \quad (A1)$$

and subsequently take $\rho_S(t)$ outside the integral. Third, according to the first assumption we can take $t-t_0 \gg \tau_c$, which gives in combination with the fact that the integrand is only nonzero on a time interval τ_c that we can replace t_0 by minus infinity. Combining everything then yields

$$i\frac{d}{dt}\rho_S(t) = (L_S - i\Gamma_M)\rho_S(t), \quad (A2)$$

with

$$\Gamma_M\sigma_S = \mathrm{Tr}_R\, L_I \int_0^\infty d\tau\, e^{-i(L_S+L_R)\tau} L_I e^{iL_S\tau}(\sigma_S\bar{\rho}_R). \quad (A3)$$

The Laplace transform of Eq. (A3) reads

$$\tilde{\rho}_S(\omega) = [i/(\omega - L_S + i\Gamma_M)]\rho_S(t_0), \quad (A4)$$

and comparison with Eq. (3.5) then shows that Γ_M is the Markovian equivalent of $\tilde{\Gamma}(\omega)$, and indeed, the frequency dependence has disappeared.

There exists an interesting relation between Γ_M and $\tilde{\Gamma}(\omega)$, which can be found as follows. In Eq. (3.9) we substitute $\exp(iL_S\tau)\sigma_S$ for σ_S and integrate the result over τ. With Eq. (A3) we then obtain

$$\Gamma_M = \int_0^\infty d\tau\, \Gamma(\tau) e^{iL_S\tau}, \quad (A5)$$

as an operator identity. Then we notice that Eq. (3.8) can be inverted as

$$\Gamma(\tau) = \frac{1}{2\pi}\int_{-\infty}^\infty d\omega\, e^{-i\omega\tau}\tilde{\Gamma}(\omega), \quad (A6)$$

for $\tau > 0$. Substitution into Eq. (A5) and performing the τ integration then leads to

$$\Gamma_M = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, \widetilde{\Gamma}(\omega)\, \frac{i}{L_S - \omega}. \qquad (A7)$$

With the projectors of Sec. VII we can write $L_S - \omega$ as

$$(L_S - \omega)\sigma_S = \sum_{ab} (\Delta_{ab} - \omega) P_a \sigma_S P_b, \qquad (A8)$$

and taking matrix elements of both sides gives

$$\langle a,\alpha | ((L_S - \omega)\sigma_S) | b,\beta \rangle = (\Delta_{ab} - \omega) \langle a,\alpha | \sigma_S | b,\beta \rangle. \qquad (A9)$$

This shows that the Liouvillian $L_S - \omega$ is diagonal with respect to the eigenstates of H_S, and that its matrix elements are $\Delta_{ab} - \omega$. Therefore, its inverse $1/(L_S - \omega)$ has corresponding matrix elements $1/(\Delta_{ab} - \omega)$, which gives the expansion of $1/(L_S - \omega)$ in projectors as

$$\frac{1}{L_S - \omega}\sigma_S = \sum_{ab} \frac{1}{\Delta_{ab} - \omega} P_a \sigma_S P_b. \qquad (A10)$$

If we insert this into Eq. (A7) and remember the general property

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega' \frac{i}{\omega - \omega'} \tilde{g}(\omega') = \tilde{g}(\omega), \qquad (A11)$$

for any Laplace transform $\tilde{g}(\omega)$, we finally obtain

$$\Gamma_M \sigma_S = \sum_{ab} \widetilde{\Gamma}(\Delta_{ab}) (P_a \sigma_S P_b). \qquad (A12)$$

Another way to derive Eq. (A12) is by substituting the expansion (7.6) for $\exp(iL_S\tau)$ into Eq. (A5) and performing the τ integration. Equation (A12) reveals that the relaxation operator in the Markov approximation effectively filters out these ω values in $\widetilde{\Gamma}(\omega)$ which are in exact resonance with the system frequencies.

[1]G. S. Agarwal, in *Quantum Statistical Theories of Spontaneous Emission and their Relation to Other Approaches, Springer Tracts in Modern Physics*, Vol. 70, edited by G. Höhler (Springer, New York, 1974).

[2]W. H. Louisell, *Quantum Statistical Properties of Radiation* (Wiley, New York, 1973).

[3]C. Cohen-Tannoudji, in *Frontiers in Laser Spectroscopy*, Proceedings of the 27th Les Houches Summer School, edited by R. Balian, S. Haroche, and S. Liberman (North-Holland, Amsterdam, 1977).

[4]S. Efrima, C. Jedrzejek, K. F. Freed, E. Hood, and H. Metiu, J. Chem. Phys. **79**, 2436 (1983).

[5]H. J. Kreuzer and Z. W. Gortel, *Physisorption Kinetics* (Springer, New York, 1986).

[6]H. F. Arnoldus, S. van Smaalen, and T. F. George, Phys. Rev. B **34**, 6902 (1986).

[7]A. Peremans, J. Darville, J. M. Gilles, and T. F. George, Phys. Rev. B **35**, 2690 (1987).

[8]R. W. Zwanzig, J. Chem. Phys. **33**, 1338 (1960).

[9]R. W. Zwanzig, in *Lectures in Theoretical Physics*, edited by W. E. Britten, B. W. Downs, and J. Downs (Interscience, New York, 1961), Vol. III, p. 106ff.

[10]C. R. Willis and R. H. Picard, Phys. Rev. A **9**, 1343 (1974).

[11]U. Fano, Phys. Rev. **131**, 259 (1963).

[12]J. Dalibard, J. Dupont-Roc, and C. Cohen-Tannoudji, J. Phys. (Paris) **43**, 1617 (1982).

[13]J. Dalibard, J. Dupont-Roc, and C. Cohen-Tannoudji, J. Phys. (Paris) **45**, 637 (1984).

[14]P. A. Apanasevich, S. Ya. Kilin, A. P. Nizovtsev, and N. S. Onishchenko, Opt. Commun. **52**, 279 (1984).

[15]M. Yamanoi and J. H. Eberly, Phys. Rev. A **34**, 1609 (1986).

[16]S. Ya. Kilin and A. P. Nizovtsev, J. Phys. B **19**, 3457 (1986).

[17]H. F. Arnoldus, S. van Smaalen, and T. F. George, to be published in Adv. Chem. Phys.

[18]Exceptions are the following: (1) G. Nienhuis, J. Phys. B **15**, 535 (1982), where τ_c equals the collision time between an atom and a perturber. [$I(\omega)$ is the fluorescence spectrum of a laser-driven atom in a bath of neutrals]; (2) H. F. Arnoldus and G. Nienhuis, J. Phys. A **19**, 1629 (1986), where τ_c is the correlation time of a stochastic process, which drives a multiplicative stochastic differential equation.

Comment on a paper by Alain J. Phares and Francis J. Wunderlich [J. Math. Phys. 26, 2491 (1985)]

John C. Wheeler
Department of Chemistry, B-014, University of California, San Diego, La Jolla, California 92093

(Received 24 March 1987; accepted for publication 15 July 1987)

The claim of Phares and Wunderlich [J. Math. Phys. **26**, 2491 (1985)] to have obtained the exact closed form analytic solution to the problem of dimers on infinite two- and three-dimensional lattices is shown to be incorrect. Their expressions do not agree with exact expansions obtained by virial and lattice counting methods.

Recently Phares and Wunderlich[1] (PW) have claimed to obtain "the exact closed-form analytic solution to the problem of dimers on infinite two-dimensional and three-dimensional lattices." They express their results in Eqs. (59)–(64) which give various thermodynamic properties as functions of the activity of a dimer x or the coverage fraction θ and a quantity $\phi(L,M)$, the "molecular freedom per dimer at close packing," that approaches a well-defined limiting value in the thermodynamic limit. Phares and Wunderlich assert in their abstract, introduction, and conclusion that this result becomes exact for infinite two- and three-dimensional lattices. Nowhere in their paper do they give rigorous or compelling arguments for this assertion. Still more recently[2] PW have expressed the expectation that the scheme developed in Ref. 1 should become exact for triangular lattices as well.

We point out here that their formulas do not agree with exact series expansions for the thermodynamic potential obtained by standard elementary counting methods even to second order in the activity x and so can not be exact. Moreover, their results do not agree with series estimates[3] for the asymptotic behavior at high coverage, and so are likely to be seriously in error in this limit. A comparison with the series results at intermediate densities suggests that their results are not particularly accurate, even as approximations, compared with fitting functions constructed from the known series.

In the limit that the activity x is small, Eq. (59) of PW can be expanded to give

$$\Gamma(x) = (2/Q)(\phi x - (3/2)(\phi x)^2 + O(\phi x)^3), \qquad (1)$$

where ϕ is the molecular freedom per dimer at close packing in the thermodynamic limit and is equal to 1.0 in one dimension (linear chain) $d = 1$, and is given according to PW, by their Eqs. (54) and (55) for two and three dimensions ($d = 2, 3$), respectively, and where $Q = 2, 4, 6$ for $d = 1, 2, 3$, respectively. In contrast, the low activity expansion on a d-dimensional hypercubic lattice is readily found to second order in activity by standard cluster integral[4,5] or lattice counting [3,6,7] methods to be

$$\Gamma = x - [(3 + 4(d-1))/2]x^2 + O(x^3). \qquad (2)$$

The numerator in the coefficient of x^2 is simply the number of ways that a second dimer can overlap with a fixed first dimer on the interior of the lattice; 3 if the dimers are parallel, and $2^2 = 4$ for each of the $d - 1$ perpendicular directions. The existence, uniqueness, and nonzero radius of convergence of this expansion are all rigorously provable.[8]

The result in Eq. (2) is in agreement with that of PW in Eq. (1) for $d = 1$, where $Q = 2$ and $\phi = 1$, but disagrees with the result of PW for $d = 2$ and 3 for *any choice of the constant* ϕ. Consequently, it must be concluded that, whatever the merits of PW's expression as an *approximation* it cannot be the exact expression for $\Gamma(x)$ in $d = 2$ or 3 in the thermodynamic limit. It follows also, by elementary thermodynamic manipulation that PW's Eqs. (60)–(64) are also not the solution of the $d = 2, 3$ dimer problem in the thermodynamic limit, contrary to the claims of the authors in their Sec. VII. From the proven analyticity[9] of the free energy in the density of dimers, it follows that Eqs. (59)–(64) of PW cannot be the exact solution at any finite activity. Additional terms have been calculated in Refs. 3, 5, and 7 and confirm that Eqs. (59)–(64) of PW are not correct for $d = 2, 3$ in the thermodynamic limit. For example, Eq. (61) of PW for the activity x can be expanded at low θ to give

$$x = \theta(2-\theta)/4\Phi(L,M)(1-\theta)^2$$
$$= \frac{1}{2\phi}\left[\sum_{n=1}^{\infty}\binom{n+1}{2}\theta^n\right]. \qquad (3)$$

Gaunt has presented exact series expansions for this quantity for several lattices. In Table II of Ref. 3 he gives 15 terms for the square (sq) lattice, 10 terms for the triangular (pt) lattice, and 12 terms for the simple cubic (sc) lattices, among others. Gaunt's variable z corresponds with PW's variable x, while Gaunt's ρ is the same as PW's and is related to θ by $\theta = Q\rho$. Making these identifications one readily verifies that Eq. (61) of Ref. 1 and Eq. (6.3) of Ref. 2 do not agree with the series of Gaunt for any value of θ.

Gaunt has analyzed his series and finds that the divergence of the activity as $\theta \rightarrow 1$ is governed by a *nonclassical* critical exponent γ. That is,

$$x(\theta) \approx A(1-\theta)^{-\gamma}, \qquad (4)$$

with $\gamma \approx 1.75$ for the square lattice and $\gamma \approx 1.95$ for the simple cubic lattice. Equation (61) of PW implies that $\gamma = 2.0$, in disagreement with the series results. This means that Eq. (61) will almost certainly give results that are much too large for $x(\theta)$ near close packing. [In this regard it is interesting to note that Gaunt estimates $\gamma \approx 2.0$ for close packed lattices like the triangular lattice, so that Eq. (59) of PW is at least consistent for that lattice in this regard. However, the

 0022-2488/87/112739-02$02.50

value of coefficient A in Eq. (4) implied by PW is 0.106, while that estimated by Gaunt is 0.149. Here the result of PW is likely to be much too small near $\theta = 1$.]

Finally, it is possible to use the exact series expansion of Gaunt to assess the reliability of Eq. (61) of PW as an approximation at intermediate densities in the thermodynamic limit. The series in Table II of Ref. 3 allows the accurate determination of $x(\theta)$ for θ up to 0.5 for the sq and sc lattices and for θ up to 0.3 for the pt lattice. For large θ they can be reliably determined by reexpressing Gaunt's series for $z(\rho)$ as

$$z(\rho) = A(\rho)(1 - Q\rho)^{-\gamma} \qquad (5)$$

with $\gamma = 1.75$ for the sq lattice, $\gamma = 1.95$ for the sc, and $\gamma = 2.0$ for the pt lattice. The resulting series for $A(\rho)$ has much smaller coefficients than those for $z(\rho)$, and when truncated even at $\rho = \rho_0 = 1/Q$ give accurate estimates for A in Eq. (5), namely $A = 0.3037, 0.1496, 0.1464$ for the sq, pt, and sc lattices, respectively, compared with Gaunt's estimates of 0.3030($\pm$4), 0.149($\pm$1), and 0.1457($\pm$1) by different methods. By examining the consistency among partial sums and Padé approximants[6] to the series for $A(\rho)$ and the variation with small changes in the exponent γ, we estimate that the truncated partial sum to $A(\rho)$ gives $z(\rho)$ through Eq. (5) to better than 1% at $\theta = 0.9$ for the sq, pt, sc lattices and rapidly become much more accurate as θ decreases. Using the approximation to $A(\rho)$ as well as the partial sums to $z(\rho)$ we have examined the accuracy of Eq. (61) of PW. For the sq lattice the percentage errors in Eq. (61) of PW at $\theta = 0.1$, 0.3, 0.5, 0.7, and 0.9 are 8.8%, 2.7%, -3.5%, -9.1%, and -5.8%, with the percentage error diverging to $+\infty$ as $\theta \to 1$. For the pt they are 22.9%, 13.3%, 2.4%, -9.4%, and -22.3% while for the sc lattice they are 18.7%, 10.0%, 0.9%, -9.2%, and -13.8%. While the values of $z(\rho)$ used here are also, of necessity, approximate they are almost surely accurate to better than a few parts per thousand even at $\theta = 0.9$ and are surely far more accurate for small values of θ.

In summary, Eqs. (59)–(64) of Ref. 1 do not become exact in the thermodynamic limit and do not have the correct behavior at either low or high coverage; Eq. (61) is not particularly accurate even at intermediate densities as an approximation to the limiting thermodynamic behavior.

[1] A. J. Phares and F. J. Wunderlich, J. Math. Phys. **26**, 2491 (1985).
[2] A. J. Phares and F. J. Wunderlich, J. Math. Phys. **27**, 1099 (1986).
[3] D. S. Gaunt, Phys. Rev. **179**, 174 (1969).
[4] J. E. Mayer and M. G. Mayer, *Statistical Mechanics* (Wiley, New York, 1940), Chap. 13.
[5] G. S. Rushbrooke, H. I. Scoins, and A. J. Wakefield, Discuss. Faraday Soc. **15**, 57 (1953).
[6] M. E. Fisher, *Lectures in Theoretical Physics* (University of Colorado, Boulder, CO, 1965), Vol. 7C, Chap. 6.
[7] J. F. Nagle, Phys. Rev. **152**, 190 (1966).
[8] D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969), Chap. 4.
[9] O. J. Heilmann and E. J. Lieb, Phys. Rev. Lett. **24**, 1412 (1970).

Canonical formalism of dissipative fields in thermo field dynamics

T. Arimitsu
Institute of Physics, University of Tsukuba, Ibaraki 305, Japan
H. Umezawa and Y. Yamanaka
Theoretical Physics Institute, Department of Physics, University of Alberta, Edmonton, Alberta T6G 2J1, Canada

(Received 30 December 1986; accepted for publication 1 July 1987)

For the stationary case the canonical formalism of thermally dissipative fields with both positive- and negative-frequency parts is constructed. This formulation enables one to follow the self-consistent renormalization scheme which creates the dissipation spontaneously. The self-interacting φ^3 model is examined as an example of the spontaneous creation of dissipation. The parameter α appearing in the thermal state conditions as well as observables independent of the choice of α are discussed.

I. INTRODUCTION

The quantum field theory provides us with a useful formalism for quantum systems with infinite degrees of freedom. A program of reformulating the nonequilibrium quantum statistical mechanics with infinite degrees of freedom in the terminology of quantum field theory has been developed by extending thermo field dynamics (TFD) which is quantum field theory with thermal degrees of freedom.[1-8] This extended TFD was shown to be equivalent to the density matrix formalism with the Liouville equation.[1] The extended TFD has so far been formulated in terms of harmonic oscillators. The purpose of this paper is to reformulate the extended TFD as a formalism for quantum field theory.

As is now widely known, TFD is built on the concepts of thermal doublets, the thermal vacuum, and the Hamiltonian.[9,10] Thus with any operator A there is associated another operator $\widetilde{A}$ which is called the tilde conjugate of A. The doublet A^μ with $A^1 = A$ and $A^2 = \widetilde{A}^\dagger$ is called the thermal doublet. The tilde conjugation rules are summarized as

$$[AB]^\sim = A^\sim B^\sim , \tag{1.1a}$$

$$[c_1 A + c_2 B]^\sim = c_1^* \widetilde{A} + c_2^* \widetilde{B} , \tag{1.1b}$$

$$[A^\dagger]^\sim = \widetilde{A}^\dagger , \tag{1.1c}$$

$$[\widetilde{A}]^\sim = \sigma A , \tag{1.1d}$$

$$|0\rangle^\sim = |0\rangle , \tag{1.1e}$$

$$\langle 0|^\sim = \langle 0| , \tag{1.1f}$$

where $|0\rangle$ and $\langle 0|$ are the thermal vacua, $\sigma = +1(-1)$ for bosonic (fermionic) A and the c_i's are c numbers.

The Hamiltonian is constructed as follows. When a system of quantum fields is given, we write the Lagrangian density $\mathscr{L}[\psi]$ from which the canonical Hamiltonian $H[\psi]$ follows through the well-known route. Then we construct $\widetilde{H}$ by applying the tilde conjugation rules to H. The Hamiltonian is then given by

$$\hat{H} = H - \widetilde{H} . \tag{1.2}$$

This statement is quite general. It covers all thermal situations, both equilibrium and nonequilibrium.

To make our consideration explicit, let us consider a nonequilibrium transition from a situation of normal conductivity to one of superconductivity. The initial situation is described by the normal quasielectron field, while the final situation is described by the superconducting quasielectron field. The intermediate situation is described by a time-dependent quasiparticle field (or a renormalized field). Thus a reasonable computational method may be the perturbative calculation in which the unperturbed representation corresponds to the quasiparticle field. Since the key mechanism in nonequilibrium phenomena is thermal dissipation, the quasiparticle field under consideration should be dissipative. In the sense that the field equations of the quasiparticles are linear and homogeneous differential equations, they can be said to be free. However, since they are dissipative, they are not really free. We have therefore called them *semifree*. Using this terminology we use semifree fields for the unperturbed representation.

It may be important to note that the dissipative effect is important even in a stationary case. Consider a situation in which the ground state is an equilibrium state. Even in such a case the excited states approach the ground state dissipatively. Then although the thermal averages of observables are independent of time, the two-point functions such as the Green's functions or correlation functions exhibit dissipative effects caused by the contributions due to the excited states. This means that the Hamiltonian for the semifree fields should contain thermally dissipative terms.

It is obvious from the above consideration that the perturbative calculation formalism requires knowledge of the general structure of the semifree field theory as its beginning. Once we know the Hamiltonian say ($\hat{H}^0$) of the semifree field, then crudely speaking $\hat{H} - \hat{H}^0$ acts as the interaction Hamiltonian. Then the Feynman diagram method tells us how to proceed in the perturbative computation. Just as the physical mass is determined by the self-consistent renormalization method that leads to mass equations, the dissipative constant is to be determined by the self-consistent renormalization method that leads to equations for dissipative coefficients. When the latter equations give rise to a nonvanishing dissipative coefficient the phenomenon is called the spontaneous creation of dissipation.[5-7] In the framework of TFD this phenomenon is expected to happen in almost all cases.

In a simple example it was shown by an exact solution.[7]

Since the Hamiltonian of a semifree system consists of the thermal doublets, the dissipative term has the form of a 2×2 thermal matrix like $\bar{a}^\mu A^{\mu\nu} a^\nu$. At first glance it may seem that the structure of the matrix A is quite arbitrary, but in fact the matrix A was found to have a very particular form. In particular, the existence of the off-diagonal elements of A, which combine tilde and nontilde operators, indicates that this dissipation is a thermal effect.

So far the study of the structure of semifree systems has been made only for harmonic oscillator-type operators. One purpose of this paper is to formulate the theory in terms of *semifree fields* instead of in terms of harmonic oscillator-type operators.

It is worth commenting on the dynamical map in TFD (i.e., the expression of Heisenberg fields in terms of certain free fields). The thermal instability giving rise to the imaginary terms in the self-energy forces us to abandon the dynamical map in terms of the usual asymptotic free fields based on the stable particle picture. In fact a no go theorem[11] states that the S matrix is trivial when the dynamical maps are expressed in terms of asymptotic free fields in TFD. Instead we express the dynamical map in TFD in terms of the semifree fields:

$$\psi \overset{w}{=} \psi[\varphi], \tag{1.3}$$

where ψ and φ are the Heisenberg field and the semifree field, respectively.[8] The symbol $\overset{w}{=}$ is the weak equality (i.e., the equality of matrix elements in reference to the Fock space of the semifree field φ). We feel that the reasons why (1.3) is possible are the negative energy of the tilde quantum and the infinite number of degrees of freedom. The creation or annihilation of a tilde quantum describes a change in the thermally excited background field and is therefore not observed as particle creation or annihilation. Intuitively speaking, the nontilde particles are acting under the influence of the thermally excited background field. This is the reason for the thermal instability of particles; they become unstable through communication with the thermal background field. The eigenvalues of $\hat{H}^0$ with imaginary dissipative part may have nothing to do with eigenvalues of $\hat{H}$ and furthermore $\hat{H}$ itself has no eigenstate in the present realization space (i.e., the Fock space of φ). The latter statement reminds us of the fact that a generator G has no eigenstates in the representation space in which the symmetry generated by G is spontaneously broken in quantum field theory.

It has been shown that there is a parameter α in TFD with the property that the physical results are independent of α. The Keldysh–Schwinger formalism[12] corresponds to the choice $\alpha = 1$, while the so-called equilibrium TFD uses $\alpha = \frac{1}{2}$.[9,10] In modern TFD for operators of harmonic oscillators the choice of α has been left undetermined. In Sec. III the structure of α transformations will be presented.

In Sec. IV the structure of the *semifree fields* will be presented. Here we restrict our consideration to stationary situations only, and we choose $\alpha = \frac{1}{2}$. The semifree field is expressed in terms of an orthonormalized complete set of wave functions which satisfy the *canonical sum rule*[8]; this situation is exactly the same as the one for the usual free field theory. In this way the semifree fields acquire the usual *canonical formalism*. The semifree field theory with *arbitrary* choice of α in *both* stationary and time-dependent situations will be presented in a separate paper.

In Sec. V we analyze a self-interacting scalar field in a stationary situation. It will be explicitly shown how the self-consistent equation for the dissipative coefficient emerges from the renormalization calculation. Althought this renormalization method has been presented only in consideration of a very simple model, the general method for choosing the renormalization point in stationary situations will be illustrated.

As a preparation for the analysis in this paper, in Sec. II, there will be given a very brief summary of the semifree oscillator theory.[5]

II. TFD IN TERMS OF SEMIFREE OSCILLATOR

We consider the oscillator variables classified by the "momentum" $\mathbf{k}$, i.e., $a(\mathbf{k})$ and $\tilde{a}(\mathbf{k})$. The thermal doublets are

$$a^\mu(\mathbf{k}):\quad a^1(\mathbf{k}) = a(\mathbf{k}),\quad a^2 = \tilde{a}^\dagger(\mathbf{k}), \tag{2.1a}$$

$$\bar{a}^\mu(\mathbf{k}):\quad \bar{a}^1(\mathbf{k}) = a^\dagger(\mathbf{k}),\quad \bar{a}^2 = -\sigma\tilde{a}(\mathbf{k}). \tag{2.1b}$$

The thermal vacuum is denoted by $|0\rangle$ and $\langle 0|$. We can write

$$\bar{a}^\mu(\mathbf{k}) = a^\dagger(\mathbf{k})\tau_\sigma \tag{2.2a}$$

with

$$\tau_\sigma = \begin{pmatrix} 1 & 0 \\ 0 & -\sigma \end{pmatrix}. \tag{2.2b}$$

We have

$$[a^\mu(\mathbf{k}),\bar{a}^\nu(\mathbf{l})]_\sigma = \delta_{\mu\nu}\delta(\mathbf{k}-\mathbf{l}) \tag{2.3}$$

with $\sigma = \pm 1$.

In this section and Sec. III, we include both the stationary and time-dependent situations in our considerations. The time evolution of operator is given by

$$a(t,\mathbf{k})^\mu = \hat{S}^{-1}(t)a(\mathbf{k})^\mu\hat{S}(t), \tag{2.4a}$$

$$\bar{a}(t,\mathbf{k})^\mu = \hat{S}^{-1}(t)\bar{a}(\mathbf{k})^\mu\hat{S}(t), \tag{2.4b}$$

with $\hat{S}(t=0) = 1$ and

$$\partial_t\hat{S}(t) = -i\hat{H}_t^0\hat{S}(t). \tag{2.5}$$

The relations in (2.4) are consistent with the tilde conjugation rules (1.1) when and only when the semifree Hamiltonian $\hat{H}_t^0$ is tildian,

$$(i\hat{H}_t^0)^\sim = i\hat{H}_t^0. \tag{2.6}$$

In the quantum field theory without thermal freedom the vacuum is empty of physical particles, implying $a(\mathbf{k})|0\rangle = 0$. When we have the thermal degree of freedom, the vacuum contains thermally excited particles. Thus $a(\mathbf{k})|0\rangle \neq 0$. It has been shown[1] that the vacuum is conditioned by the thermal state conditions which read as

$$a(t,\mathbf{k})^1|0\rangle = f^\alpha(t,\mathbf{k})a(t,\mathbf{k})^2|0\rangle, \tag{2.7a}$$

$$\langle 0|\bar{a}(t,\mathbf{k})^1 = -\sigma f^{1-\alpha}(t,\mathbf{k})\langle 0|\bar{a}(t,\mathbf{k})^2, \tag{2.7b}$$

with a real and positive function $f(t,\mathbf{k})$ and $0 \leqslant \alpha \leqslant 1$. The parameter α may depend on time.

Considering the tilde conjugation rules (1.1) we have from (2.7)

$$\bar{a}(t,\mathbf{k})^2|0\rangle = -f^{\alpha}(t,\mathbf{k})\bar{a}(t,\mathbf{k})^1|0\rangle \,, \tag{2.8a}$$

$$\langle 0|a(t,\mathbf{k})^2 = \sigma f^{1-\alpha}(t,\mathbf{k})\langle 0|a(t,\mathbf{k})^1 \,. \tag{2.8b}$$

The relations (2.7) and (2.8) give the complete set of $[\alpha]$-representation thermal state conditions for the semifree oscillators. In deriving (2.8) we have used the assumption that $f(t,\mathbf{k})$ is real which is equivalent to considering only systems with real particle number

$$n(t,\mathbf{k}) = \langle 0|\bar{a}(t,\mathbf{k})^1 a(t,\mathbf{k})^1|0\rangle = f(t,\mathbf{k})/[1-\sigma f(t,\mathbf{k})] \,, \tag{2.9}$$

where (2.7) and (2.8) have been used. When $f(t,\mathbf{k})$ is independent of t [i.e., $f(t,\mathbf{k}) = f(\mathbf{k})$], we state that the situation is stationary.

Since $\hat{S}(t)$ is not necessarily unitary, we used the symbol †† instead of †. Note that the nonequilibrium TFD presented in Refs. 1 and 2 used the $[\alpha = 1]$ representation, while the equilibrium TFD in Refs. 9 and 10 employed the $[\alpha = \frac{1}{2}]$ representation. The $[\alpha = 0]$ representation corresponds to *the mirror space*.[1] The physical quantities, the detailed definition of which will be given in the next section, are independent of the choice of α.

It follows from the thermal state conditions that $\hat{H}_t^0$ has the following structure:

$$\hat{H}_t^0 = \int d^3k\, \bar{a}(\mathbf{k})^{\mu}[\omega(t,\mathbf{k})\delta^{\mu\nu} - i\kappa(t,\mathbf{k})A(t,\mathbf{k})^{\mu\nu} + \sigma\{\partial_t n(t,\mathbf{k})\}\tau(t,\mathbf{k})^{\mu\nu} + \{\partial_t \ln f(t,\mathbf{k})^{(1-\alpha)/2}\}\tau_3^{\mu\nu}]a(\mathbf{k})^{\nu} \,. \tag{2.10}$$

Here $\kappa(t,\mathbf{k})$ is the dissipative coefficient. The matrices A and τ are

$$A(t,\mathbf{k}) = \begin{bmatrix} 1+2\sigma n(t,\mathbf{k}) & -2f^{\alpha-1}(t,\mathbf{k})n(t,\mathbf{k}) \\ 2\sigma f^{1-\alpha}(t,\mathbf{k})[1+\sigma n(t,\mathbf{k})] & -[1+2\sigma n(t,\mathbf{k})] \end{bmatrix}, \tag{2.11}$$

$$\tau(t,\mathbf{k}) = \begin{bmatrix} 1 & -\sigma f^{\alpha-1}(t,\mathbf{k}) \\ \sigma f^{1-\alpha}(t,\mathbf{k}) & -1 \end{bmatrix}, \tag{2.12}$$

and the τ_i's $(i = 1,2,3)$ are Pauli matrices.

The thermal state conditions (2.7) and (2.8) have the form

$$\gamma(t,\mathbf{k})^1|0\rangle = \bar{\gamma}(t,\mathbf{k})^2|0\rangle = 0 \,, \tag{2.13a}$$

$$\langle 0|\gamma(t,\mathbf{k})^2 = \langle 0|\bar{\gamma}(t,\mathbf{k})^1 = 0 \,, \tag{2.13b}$$

where

$$\gamma(t,\mathbf{k})^{\mu} = B(t,\mathbf{k})^{\mu\nu}a(t,\mathbf{k})^{\nu} \,, \tag{2.14a}$$

$$\bar{\gamma}(t,\mathbf{k})^{\mu} = \bar{a}(t,\mathbf{k})^{\nu}B^{-1}(t,\mathbf{k})^{\nu\mu} \,, \tag{2.14b}$$

with

$$B(t,\mathbf{k}) = [1-\sigma f(t,\mathbf{k})]^{-1/2} \times \begin{bmatrix} 1 & -f^{\alpha}(t,\mathbf{k}) \\ -\sigma f^{1-\alpha}(t,\mathbf{k}) & 1 \end{bmatrix}. \tag{2.15}$$

In defining γ^{μ} and $\bar{\gamma}^{\mu}$, we have imposed the condition that $\det B = 1$. Equations (2.14) lead to

$$[\gamma^{\mu}(t,\mathbf{k}),\bar{\gamma}^{\nu}(t,\mathbf{l})]_{\sigma} = \delta_{\mu\nu}\delta(\mathbf{k}-\mathbf{l}) \,. \tag{2.16}$$

The equations in (2.13) indicate that $\gamma(t,\mathbf{k})^1$ and $\bar{\gamma}(t,\mathbf{k})^2$ are the annihilation operators while $\gamma(t,\mathbf{k})^2$ and $\bar{\gamma}(t,\mathbf{k})^1$ are the creation operators. We can prove

$$B^{-1}(t,\mathbf{k}) = \tau_3 B(t,\mathbf{k})\tau_3 \,. \tag{2.17}$$

It has been shown that $\gamma(t,\mathbf{k})^{\mu}$ and $\bar{\gamma}(t,\mathbf{k})^{\mu}$ are the eigenfunctions of the form

$$\gamma(t,\mathbf{k})^{\mu} = \left[W(t,0,\mathbf{k})\exp\int_0^t d\tau\{-i\omega(\tau,\mathbf{k}) - \tau_3\kappa(\tau,\mathbf{k})\}\right]^{\mu\nu}\gamma(t=0,\mathbf{k})^{\nu} \,, \tag{2.18a}$$

$$\bar{\gamma}(t,\mathbf{k})^{\mu} = \bar{\gamma}(t=0,k)^{\nu}\left[\exp\int_0^t d\tau\{i\omega(\tau,\mathbf{k}) + \tau_3\kappa(\tau,\mathbf{k})\}W^{-1}(t,0,\mathbf{k})\right]^{\nu\mu} , \tag{2.18b}$$

with

$$W(t,s,\mathbf{k}) = \begin{bmatrix} z(t,s,\mathbf{k}) & 0 \\ 0 & z^{-1}(t,s,\mathbf{k}) \end{bmatrix}, \tag{2.19}$$

where

$$z(t,s,\mathbf{k}) = \left[\frac{n(s,\mathbf{k})}{n(t,\mathbf{k})}\right]^{(1-\alpha)/2}\left[\frac{1+\sigma n(s,\mathbf{k})}{1+\sigma n(t,\mathbf{k})}\right]^{\alpha/2} \tag{2.20}$$

$$= z^{-1}(s,t,\mathbf{k}) \,. \tag{2.21}$$

This leads to the commutation relation for arbitrary times as

$$[\gamma(t,\mathbf{k})^{\mu},\bar{\gamma}(s,\mathbf{l})^{\nu}]_{\sigma} = \left[W(t,s,\mathbf{k})\exp\int_s^t d\tau\{-i\omega(\tau,\mathbf{k}) - \tau_3\kappa(\tau,\mathbf{k})\}\right]^{\mu\nu}\delta(\mathbf{k}-\mathbf{l}) \,. \tag{2.22}$$

The representation space in TFD (called the thermal space) is the vector space spanned by the set of bra and ket state vectors which are generated, respectively, by cyclic operations of the annihilation operators γ^1 and $\bar{\gamma}^2$ on $\langle 0|$, and of the creation operators γ^2 and $\bar{\gamma}^1$ on $|0\rangle$.

Rewriting physical operators $a(t,\mathbf{k})^{\mu}$ and $\bar{a}(t,\mathbf{k})^{\mu}$ in terms of the operators $\gamma(t,\mathbf{k})^{\mu}$ and $\bar{\gamma}(t,\mathbf{k})^{\mu}$, and using the commutation relation (2.22), we can rewrite any product as a sum of normal products, with γ^1, $\bar{\gamma}^2$ to the right of γ^2, $\bar{\gamma}^1$. This leads to a Wick-type formula, which in turn leads to Feynman-type diagrams for multipoint functions in the renormalized interaction representation. The internal line in the Feynman-type diagrams is the unperturbed two-point function, the calculation of which will be given in the following.

The calculation of the unperturbed causal two-point function,

$$G(t,s,\mathbf{k})^{\mu\nu}\delta(\mathbf{k}-\mathbf{l}) = -i\langle 0|T[a(t,\mathbf{k})^{\mu}\bar{a}(s,\mathbf{l})^{\nu}]|0\rangle , \tag{2.23}$$

can be made by means of the method of the Wick-type formula. The result has been obtained. It is

$$G(t,s,\mathbf{k})^{\mu\nu} = [B(t,\mathbf{k})\mathscr{G}(t,s,\mathbf{k})B^{-1}(s,\mathbf{k})]^{\mu\nu}, \tag{2.24}$$

where

$$\mathscr{G}(t,s,\mathbf{k})^{\mu\nu}\delta(\mathbf{k}-l) = -i\langle 0|T[\gamma(t,\mathbf{k})^{\mu}\bar{\gamma}(s,l)^{\nu}]|0\rangle , \tag{2.25}$$

whose elements are explicitly given by

$$\mathscr{G}(t,s,\mathbf{k})^{11} = z(t,s,\mathbf{k})G^{r}(t,s,\mathbf{k}) , \tag{2.26a}$$

$$\mathscr{G}(t,s,\mathbf{k})^{22} = z(s,t,\mathbf{k})G^{a}(t,s,\mathbf{k}) , \tag{2.26b}$$

and $\mathscr{G}(t,s,\mathbf{k})^{12} = \mathscr{G}(t,s,\mathbf{k})^{21} = 0$, with

$$G^{r}(t,s,\mathbf{k}) = -i\theta(t-s)\exp\left[\int_s^t d\tau \times\{-i\omega(\tau,\mathbf{k})-\kappa(\tau,\mathbf{k})\}\right], \tag{2.27a}$$

$$G^{a}(t,s,\mathbf{k}) = i\theta(s-t)\exp\left[\int_s^t d\tau \times\{-i\omega(\tau,\mathbf{k})+\kappa(\tau,\mathbf{k})\}\right]. \tag{2.27b}$$

III. THE α TRANSFORMATION AND OBSERVABLES

As is seen from the arguments in the previous section, we have infinite ways of representing a thermal situation through the thermal state conditions. This freedom was indicated by α in the thermal state conditions (2.7) and (2.8). (In the density operator formalism[13] the α freedom arises from the trace formula $\mathrm{Tr}[\rho A] = \mathrm{Tr}[\rho^{1-\alpha}A\rho^{\alpha}]$ with any operator A and the Liouville equation of the form $\dot{\rho}^{\alpha} = -i[\rho^{\alpha},H]$.) According to (2.4), (2.5), and (2.10), the parameters and operators appearing in the free dissipative Hamiltonian also depend on α, and therefore $\hat{H}^{0}(t)$,

$$\hat{H}^{0}(t) = \hat{S}^{-1}(t)\hat{H}^{0}_{t}\hat{S}(t) \tag{3.1a}$$

should be denoted by $\hat{H}^{0}_{\alpha}(t)$:

$$\hat{H}^{0}_{\alpha}(t) = \hat{H}^{0}(a_{\alpha}(t),\bar{a}_{\alpha}(t);\alpha) . \tag{3.1b}$$

Thus both the thermal state conditions and the free dissipative Hamiltonian depend on α. The fact that the thermal state conditions depend on α implies that the thermal vacuum also depends on α. Thus we should write $|0_{\alpha}\rangle$ and $\langle 0_{\alpha}|$. However, in the following sections, the suffix α in the Hamiltonian and the thermal vacua will be mostly omitted.

The observable results, however, should be independent of α. Therefore it is convenient to find the transformation which changes α as $\alpha\to\alpha'$ in order to identify the observable operators.

Let $T(0)$ denote the operator inducing the following transformation:

$$|0_{\alpha'}\rangle = T^{-1}(0)|0_{\alpha}\rangle , \quad \langle 0_{\alpha'}| = \langle 0_{\alpha}|T(0) . \tag{3.2}$$

Without loss of generality we can prepare all of the operators in such a way that they become independent of α at $t=0$:

$$a(0)^{\mu}\equiv a_{\alpha}(0)^{\mu} = a_{\alpha'}(0)^{\mu} ,$$
$$\bar{a}(0)^{\mu}\equiv\bar{a}_{\alpha}(0)^{\mu} = \bar{a}_{\alpha'}(0)^{\mu} . \tag{3.3}$$

Then $T(0)$ can be easily obtained from the thermal state conditions at $t=0$ as

$$T(0) = \exp[-\tfrac{1}{2}\delta\alpha(0)\ln f(0)\bar{a}(0)\tau_3 a(0)] , \tag{3.4}$$

with $\delta\alpha(t) = \alpha'(t)-\alpha(t)$.

Since

$$A^{(r)}_{\alpha}(t)^{\mu\nu} \equiv\langle 0_{\alpha}|a_{\alpha}(t)^{\mu}\bar{a}_{\alpha}(t)^{\nu}|0_{\alpha}\rangle \tag{3.5a}$$

$$= \begin{bmatrix} 1+\sigma n(t) & -f^{\alpha-1}(t)n(t) \\ \sigma f^{1-\alpha}(t)[1+\sigma n(t)] & -\sigma n(t)\end{bmatrix}, \tag{3.5b}$$

we can relate $A^{(r)}_{\alpha'}(t)$ to $A^{(r)}_{\alpha}(t)$ as follows:

$$A^{(r)}_{\alpha'}(t) = W(t)A^{(r)}_{\alpha}(t)W^{-1}(t) . \tag{3.6}$$

Here

$$W(t) = \exp[\tfrac{1}{2}\tau_3\delta\alpha(t)\ln f(t)] . \tag{3.7}$$

This shows that

$$a_{\alpha'}(t)^{\mu} = T^{-1}(0)W(t)^{\mu\nu}a_{\alpha}(t)^{\nu}T(0) ,$$
$$\bar{a}_{\alpha'}(t)^{\mu} = T^{-1}(0)\bar{a}_{\alpha}(t)^{\nu}W^{-1}(t)^{\nu\mu}T(0) , \tag{3.8}$$

which become

$$a_{\alpha'}(t)^{\mu} = \mathscr{T}(t)a_{\alpha}(t)^{\mu}\mathscr{T}^{-1}(t) ,$$
$$\bar{a}_{\alpha'}(t)^{\mu} = \mathscr{T}(t)\bar{a}_{\alpha}(t)^{\mu}\mathscr{T}^{-1}(t) . \tag{3.9}$$

Here

$$\mathscr{T}(t) = T^{-1}(0)T(t) \tag{3.10}$$

with

$$T(t) = \exp[-\tfrac{1}{2}\delta\alpha(t)\ln f(t)\bar{a}_{\alpha}(t)\tau_3 a_{\alpha}(t)] . \tag{3.11}$$

In deriving (3.9), use was made of the relations

$$T(t)a_{\alpha}(t)^{\mu}T^{-1}(t) = W(t)^{\mu\nu}a_{\alpha}(t)^{\nu} ,$$
$$T(t)\bar{a}_{\alpha}(t)^{\mu}T^{-1}(t) = \bar{a}_{\alpha}(t)^{\nu}W^{-1}(t)^{\nu\mu} . \tag{3.12}$$

Note that

$$\mathscr{T}(0) = 1 . \tag{3.13}$$

The $\mathscr{T}(t)$ transformation changes the Hamiltonian as

$$\hat{H}^{0}_{\alpha'}(t) = \mathscr{T}(t)\hat{H}^{0}_{\alpha}(t)\mathscr{T}^{-1}(t) - i\dot{\mathscr{T}}(t)\mathscr{T}^{-1}(t) . \tag{3.14}$$

A calculation shows that this changes only the explicit α in $\hat{H}^{0}_{\alpha}(t)$ in (2.10) as $\alpha\to\alpha'$. Thus $\delta\omega$ and κ are independent of α.

An operator Q_{α} is said to be an observable, when and only when its vacuum expectation value is independent of α:

$$\langle 0_{\alpha}|Q_{\alpha}|0_{\alpha}\rangle = \langle 0_{\alpha'}|Q_{\alpha'}|0_{\alpha'}\rangle . \tag{3.15}$$

For our purpose it is sufficient to consider the Q_{α} of the form

$$Q_{\alpha}(t_1,\ldots,t_m) = Q_{\alpha 1}(t_1)Q_{\alpha 2}(t_2)\cdots Q_{\alpha m}(t_m) , \tag{3.16}$$

where each of $Q_{\alpha i}$ ($i=1,\ldots,m$) stands for any one of a_{α}, $a^{\dagger\dagger}_{\alpha}$, $\tilde{a}_{\alpha}$, and $\tilde{a}^{\dagger\dagger}_{\alpha}$. This is because the most general form of operators is a linear combination of the terms of the form in (3.16).

It follows from (3.9) that

$$\langle 0_{\alpha'}|Q_{\alpha'}|0_{\alpha'}\rangle = e^{x}\langle 0_{\alpha}|Q_{\alpha}|0_{\alpha}\rangle , \tag{3.17}$$

with

$$x=\frac{1}{2}\sum_{i=1}^{m}\epsilon_i \ln\left\{\frac{f(t_i)^{\delta\alpha(t_i)}}{f(0)^{\delta\alpha(0)}}\right\}, \tag{3.18}$$

$$\epsilon_i=\begin{cases}1, & \text{for } a_\alpha \text{ and } \tilde{a}_\alpha,\\ -1, & \text{for } a_\alpha^{\dagger\dagger} \text{ and } \tilde{a}_\alpha^{\dagger\dagger}.\end{cases} \tag{3.19}$$

Comparing (3.17) and (3.15), we see that

$$x=0 \tag{3.20}$$

is necessary and sufficient for Q_α to be an observable. In using (3.20), it is very important that α can vary in time.

As is seen from (3.18) and (3.20), any observable Q_α is a linear sum of products of the following operators:

$$\begin{aligned} &N_\alpha(t)=a_\alpha^{\dagger\dagger}(t)a_\alpha(t), \quad N_\alpha(t)=\tilde{a}_\alpha^{\dagger\dagger}(t)\tilde{a}_\alpha(t),\\ &M_\alpha(t)=a_\alpha^{\dagger\dagger}(t)\tilde{a}_\alpha(t), \quad \widetilde{M}_\alpha(t)=\sigma\tilde{a}_\alpha^{\dagger\dagger}(t)a_\alpha(t). \end{aligned} \tag{3.21}$$

Note that the operators such as $a_\alpha^{\dagger\dagger}(t_1)a_\alpha(t_2)$ with $t_1\neq t_2$ do not satisfy condition (3.20) and, therefore, are not observables.

We can summarize the above results by the statement that the observables are represented by the operators which are invariant under the following *time-local* dilatation:

$$\begin{aligned} &a_\alpha(t)\to e^{\theta(t)}a_\alpha(t), \quad \tilde{a}_\alpha(t)\to e^{\theta(t)}\tilde{a}_\alpha(t),\\ &a_\alpha^{\dagger\dagger}(t)\to e^{-\theta(t)}a_\alpha^{\dagger\dagger}(t), \quad \tilde{a}_\alpha^{\dagger\dagger}(t)\to e^{-\theta(t)}\tilde{a}_\alpha^{\dagger\dagger}(t). \end{aligned} \tag{3.22}$$

IV. GENERAL FORMALISM FOR SEMIFREE FIELD DIVISOR

In this section, we will consider properties of the semifree field for type II in the stationary case. This time-independent situation corresponds to the long time limit in the statistical mechanical argument. We will use the *[$\alpha=\frac{1}{2}$] representation* because the general formalism for type II semifree fields is most easily constructed in this representation. Here the type II field means those fields which carry both the particle (i.e., positive-frequency wave function) and antiparticles (i.e., negative-frequency wave functions). The type I semifree field consisting of the particle only is discussed in Ref. 4.

According to (2.10) with $\alpha=\frac{1}{2}$, the Hamiltonian for the type II semifree field in a stationary situation is

$$\begin{aligned}\widehat{H}^0=\int d^3k\,\{&\bar{a}(\mathbf{k})^\mu[\omega(\mathbf{k})\delta^{\mu\nu}-i\kappa(\mathbf{k})A(\mathbf{k})^{\mu\nu}]a(\mathbf{k})^\nu\\ &+\sigma\bar{b}^{\dagger}(\mathbf{k})^\mu[\omega(\mathbf{k})\delta^{\mu\nu}-i\kappa(\mathbf{k})A^T(\mathbf{k})^{\mu\nu}]b^{\dagger}(\mathbf{k})^\nu\},\end{aligned} \tag{4.1}$$

where the matrix A is defined by (2.11). We have used the thermal doublet notations both for particle $(a^\mu,\bar{a}^\mu)$ [cf. (2.1)] and antiparticle $(b^{\dagger\mu},\bar{b}^{\dagger\mu})$,

$$b^{\dagger}(\mathbf{k})^\mu:\ b^{\dagger}(\mathbf{k})^1=b^{\dagger}(\mathbf{k}),\quad b^{\dagger}(\mathbf{k})^2=\tilde{b}(\mathbf{k}), \tag{4.2a}$$

$$\bar{b}^{\dagger}(\mathbf{k})^\mu:\ \bar{b}^{\dagger}(\mathbf{k})^1=b(\mathbf{k}),\quad \bar{b}^{\dagger}(\mathbf{k})^2=-\sigma\tilde{b}^{\dagger}(\mathbf{k}). \tag{4.2b}$$

Note that the Hamiltonian and the thermal state conditions for antiparticle are obtained by the replacement $a(\mathbf{k})\to b(\mathbf{k})$ together with $\mathbf{k}\to-\mathbf{k}$ in those for particle.

In writing (4.1), we have assumed the symmetry of the particle and antiparticle (i.e., they have the same energy spectrum ω, damping parameter κ, and particle distribution n), and the isotropy of the system (i.e., the quantities ω, κ, and n are dependent on only the magnitude of $\mathbf{k}$). Note the property of the matrix A,

$$\tau_\sigma A^T\tau_\sigma=A, \tag{4.3}$$

when $\alpha=\frac{1}{2}$, where A^T indicates the transpose of A and τ_σ is given in (2.2b).

The Hamiltonian $\widehat{H}^0$ leads to the following equations of motion for the particle and antiparticle:

$$[i\partial^{+}_{t,\kappa}(\mathbf{k})^{\mu\nu}-\omega(\mathbf{k})\delta^{\mu\nu}]a(t,\mathbf{k})^\nu=0, \tag{4.4a}$$

$$\bar{a}(t,\mathbf{k})^\mu[-i\overleftarrow{\partial}^{+}_{t,-\kappa}(\mathbf{k})^{\nu\mu}-\omega(\mathbf{k})\delta^{\nu\mu}]=0, \tag{4.4b}$$

and

$$[i\partial^{-}_{t,\kappa}(\mathbf{k})^{\mu\nu}+\omega(\mathbf{k})\delta^{\mu\nu}]b^{\dagger\dagger}(t,\mathbf{k})^\nu=0, \tag{4.5a}$$

$$\bar{b}^{\dagger\dagger}(t,\mathbf{k})^\nu[-i\overleftarrow{\partial}^{-}_{t,-\kappa}(\mathbf{k})^{\nu\mu}+\omega(\mathbf{k})\delta^{\nu\mu}]=0, \tag{4.5b}$$

respectively, where

$$\partial^{\pm}_{t,\kappa}(\mathbf{k})^{\mu\nu}=\partial_t\delta^{\mu\nu}\pm\kappa(\mathbf{k})A(\mathbf{k})^{\mu\nu}. \tag{4.6}$$

Equations (4.5) and (4.6) are solved to give

$$a(t,\mathbf{k})^\mu=e^{-i\omega(\mathbf{k})t}U_\kappa(t,\mathbf{k})^{\mu\nu}a(\mathbf{k})^\nu, \tag{4.7a}$$

$$\bar{a}(t,\mathbf{k})^\mu=\bar{a}(\mathbf{k})^\nu U_{-\kappa}(t,\mathbf{k})^{\nu\mu}e^{i\omega(\mathbf{k})t}, \tag{4.7b}$$

and

$$b^{\dagger\dagger}(t,\mathbf{k})^\mu=e^{i\omega(\mathbf{k})t}U_{-\kappa}(t,\mathbf{k})^{\mu\nu}b^{\dagger}(\mathbf{k})^\nu, \tag{4.8a}$$

$$\bar{b}^{\dagger\dagger}(t,\mathbf{k})^\mu=\bar{b}^{\dagger}(\mathbf{k})^\nu U_\kappa(t,\mathbf{k})^{\nu\mu}e^{-i\omega(\mathbf{k})t}, \tag{4.8b}$$

respectively, where

$$\begin{aligned}U_\kappa(t,\mathbf{k})&=\exp[-\kappa(\mathbf{k})A(\mathbf{k})t]\\ &=A^{(r)}(\mathbf{k})e^{-\kappa(\mathbf{k})t}-A^{(a)}(\mathbf{k})e^{\kappa(\mathbf{k})t},\end{aligned} \tag{4.9}$$

where the matrix $A^{(r)}$ is defined in (3.5) and $A^{(a)}$ is given by

$$A^{(a)}(\mathbf{k})^{\mu\nu}=\langle 0|\bar{a}(\mathbf{k})^\nu a(\mathbf{k})^\mu|0\rangle=\begin{bmatrix}\sigma n(\mathbf{k}) & -\sqrt{n(\mathbf{k})[1+\sigma n(\mathbf{k})]}\\ \sigma\sqrt{n(\mathbf{k})[1+\sigma n(\mathbf{k})]} & -[1+\sigma n(\mathbf{k})]\end{bmatrix}^{\mu\nu}, \tag{4.10}$$

and has characteristics

$$[A^{(r)}(\mathbf{k})]^2=A^{(r)}(\mathbf{k}), \tag{4.11a}$$

$$[A^{(a)}(\mathbf{k})]^2=-A^{(a)}(\mathbf{k}), \tag{4.11b}$$

$$A^{(r)}(\mathbf{k})A^{(a)}(\mathbf{k})=A^{(a)}(\mathbf{k})A^{(r)}(\mathbf{k})=0, \tag{4.11c}$$

$$A^{(r)}(\mathbf{k})^{\mu\nu}-A^{(a)}(\mathbf{k})^{\mu\nu}=\delta^{\mu\nu}, \tag{4.11d}$$

$$A^{(r)}(\mathbf{k})^{\mu\nu}+A^{(a)}(\mathbf{k})^{\mu\nu}=A(\mathbf{k})^{\mu\nu}. \tag{4.11e}$$

From (4.4) and (4.5), we see that

$$-\{\partial_t^2+[\omega(\mathbf{k})-i\kappa(\mathbf{k})A(\mathbf{k})]^2\}^{\mu\nu}a(t,\mathbf{k})^\nu=0\,, \tag{4.12}$$

$$-\{\partial_t^2+[\omega(\mathbf{k})-i\kappa(\mathbf{k})A(\mathbf{k})]^2\}^{\mu\nu}b^{\dagger\dagger}(t,\mathbf{k})^\nu=0\,. \tag{4.13}$$

A. The semifree field equation for a physical field

We consider a semifree field $\varphi(x)$ of type II which satisfies the field equation

$$\lambda_\kappa(\partial)^{\mu\nu}\varphi(x)^\nu=0\,, \tag{4.14}$$

where $\partial=(\nabla,\partial_t)$. The field equation is called a type II equation when it can be reduced to the eigenvalue equation

$$\{\partial_t^2+[\omega(-i\nabla)-i\kappa(-i\nabla)A(-i\nabla)]^2\}^{\mu\nu}\varphi(x)^\nu=0\,. \tag{4.15}$$

This dissipative free field equation (the equation for a semifree field) is the extension of the nondissipative free field $[\partial_t^2+\omega^2(-i\nabla)]\varphi=0$. Note that the form of the dissipative term is not arbitrary; it should be proportional to the matrix A.

B. The divisor

For the semifree field, the divisor $d_\kappa(\partial)$ is defined by

$$\begin{aligned}\lambda_\kappa(\partial)^{\mu\delta}d_\kappa(\partial)^{\delta\nu}&=d_\kappa(\partial)^{\mu\delta}\lambda_\kappa(\partial)^{\delta\nu}\\&=-\{\partial_t^2+[\omega(-i\nabla)-i\kappa(-i\nabla)A(-i\nabla)]^2\}^{\mu\nu}.\end{aligned} \tag{4.16}$$

If we denote by $\Delta_G(x,y)$ any Green's functions of (4.15),

$$\begin{aligned}-\{\partial_t^2+[\omega(-i\nabla)-i\kappa(-i\nabla)A(-i\nabla)]^2\}^{\mu\delta}\Delta_G(x,y)^{\delta\nu}\\=\delta^{\mu\nu}\delta(\mathbf{x}-\mathbf{y})\delta(t-s)\,,\end{aligned} \tag{4.17}$$

with $y=(\mathbf{y},s)$ and $\nabla=\nabla_x$, $d_\kappa(\partial)\Delta_G(x,y)$ is a Green's function for (4.14),

$$\lambda_\kappa(\partial)^{\mu\delta}[d_\kappa(\partial)\Delta_G(x,y)]^{\delta\nu}=\delta^{\mu\nu}\delta(\mathbf{x}-\mathbf{y})\delta(t-s)\,. \tag{4.18}$$

C. The wave equation

In the usual quantum field theory without thermal degrees of freedom, the consideration of free fields begins with finding $u(x,\mathbf{k})a(\mathbf{k})$ for particles and $v(x,\mathbf{k})b^\dagger(\mathbf{k})$ for antiparticles. Here $u(x,\mathbf{k})$ and $v(x,\mathbf{k})$ are the wave functions of positive and negative frequencies, respectively, and x means $(\mathbf{x},t)$. These wave functions are usually determined by the free field equations. On the other hand, when the unperturbed Hamiltonian is known, they can also be obtained from the Hamiltonian. The Hamiltonian determines $a(t,\mathbf{k})$ and $b(t,\mathbf{k})$, and then the wave functions $u(x,\mathbf{k})$ and $v(x,\mathbf{k})$ are obtained through relations such as $a(t,\mathbf{k})u(\mathbf{x},\mathbf{k})=a(\mathbf{k})u(x,\mathbf{k})$, etc. In the case under consideration, we follow the latter method because the unperturbed Hamiltonian $\hat{H}^0$, (4.1), is known.

Through the relations

$$u(\mathbf{x},\mathbf{k})^{\mu\nu}a(t,\mathbf{k})^\nu=u(x,\mathbf{k})^{\mu\nu}a(\mathbf{k})^\nu\,, \tag{4.19a}$$

$$v(\mathbf{x},\mathbf{k})^{\mu\nu}b^{\dagger\dagger}(t,\mathbf{k})^\nu=v(x,\mathbf{k})^{\mu\nu}b^\dagger(\mathbf{k})^\nu\,, \tag{4.19b}$$

with

$$u(\mathbf{x},\mathbf{k})^{\mu\nu}=u(\mathbf{k})^{\mu\nu}e^{i\mathbf{k}\cdot\mathbf{x}}\,, \tag{4.20a}$$

$$v(\mathbf{x},\mathbf{k})^{\mu\nu}=v(\mathbf{k})^{\mu\nu}e^{-i\mathbf{k}\cdot\mathbf{x}}\,, \tag{4.20b}$$

we have

$$u(x,\mathbf{k})^{\mu\nu}=e^{-i\omega(\mathbf{k})t+i\mathbf{k}\cdot\mathbf{x}}U_\kappa(t,\mathbf{k})^{\mu\delta}u(\mathbf{k})^{\delta\nu}\,, \tag{4.21a}$$

$$v(x,\mathbf{k})^{\mu\nu}=e^{i\omega(\mathbf{k})t-i\mathbf{k}\cdot\mathbf{x}}U_{-\kappa}(t,\mathbf{k})^{\mu\delta}v(\mathbf{k})^{\delta\nu}\,. \tag{4.21b}$$

Here (4.7) and (4.8) have been used. We see that $u(x,\mathbf{k})$ and $v(x,\mathbf{k})$ satisfy

$$[i\partial_{t,\kappa}^+(-i\nabla)^{\mu\delta}-\omega(-i\nabla)\delta^{\mu\delta}]u(x,\mathbf{k})^{\delta\nu}=0, \tag{4.22a}$$

$$[i\partial_{t,\kappa}^-(-i\nabla)^{\mu\delta}+\omega(-i\nabla)\delta^{\mu\delta}]v(x,\mathbf{k})^{\delta\nu}=0. \tag{4.22b}$$

The physical field $\varphi(x)$ is expressed as

$$\varphi(x)^\mu=\int d^3k\,[u(x,\mathbf{k})^{\mu\nu}a(\mathbf{k})^\nu+v(x,\mathbf{k})^{\mu\nu}b^\dagger(\mathbf{k})^\nu]\,. \tag{4.23}$$

Then substitution of (4.23) into the field equation (4.14) gives us wave equations

$$\lambda_\kappa(\partial)^{\mu\delta}u(x,\mathbf{k})^{\delta\nu}=0\,, \tag{4.24a}$$

$$\lambda_\kappa(\partial)^{\mu\delta}v(x,\mathbf{k})^{\delta\nu}=0\,. \tag{4.24b}$$

D. The Hermitization matrix

As we have assumed that the differential operator $\lambda_\kappa(\partial)$ depends on the thermal degrees of freedom (i.e., the superscripts) only through $i\kappa(-i\nabla)A(-i\nabla)$, we can easily see that it has the property

$$\lambda_\kappa(-\partial)\tau_2^\dagger=\tau_2\lambda_\kappa(\partial)\,. \tag{4.25}$$

In deriving (4.25), we have used the characteristic

$$A^T(-i\nabla)=-\tau_2A(-i\nabla)\tau_2^{-1}\,. \tag{4.26}$$

In the following, we drop the superscript unless it is needed. Equation (4.25) tells us that the matrix τ_2 is the Hermitization matrix with respect to the thermal degrees of freedom.

The property (4.25) and the wave equations (4.24) give us

$$\bar{u}(x,\mathbf{k})\lambda_\kappa(-\overleftarrow{\partial})=0\,, \tag{4.27a}$$

$$\bar{v}(x,\mathbf{k})\lambda_\kappa(-\overleftarrow{\partial})=0\,, \tag{4.27b}$$

where

$$\bar{u}(x,\mathbf{k})=\tau_2u^\dagger(x,\mathbf{k})\tau_2^{-1}\,, \tag{4.28a}$$

$$\bar{v}(x,\mathbf{k})=\tau_2v^\dagger(x,\mathbf{k})\tau_2^{-1}\,. \tag{4.28b}$$

Substituting (4.21) into (4.28), we have

$$\bar{u}(x,\mathbf{k})^{\mu\nu}=\bar{u}(\mathbf{k})^{\mu\delta}U_{-\kappa}(t,\mathbf{k})^{\delta\nu}e^{i\omega(\mathbf{k})t-i\mathbf{k}\cdot\mathbf{x}}\,, \tag{4.29a}$$

$$\bar{v}(x,\mathbf{k})^{\mu\nu}=\bar{v}(\mathbf{k})^{\mu\delta}U_\kappa(t,\mathbf{k})^{\delta\nu}e^{-i\omega(\mathbf{k})t+i\mathbf{k}\cdot\mathbf{x}}\,, \tag{4.29b}$$

where we have used the relation

$$U_\kappa^T(t,\mathbf{k})=\tau_2U_{-\kappa}(t,\mathbf{k})\tau_2^{-1}\,, \tag{4.30}$$

which can be obtained from (4.26). Note that the relations

$$\bar{a}(t,\mathbf{k})\bar{u}(\mathbf{x},\mathbf{k})=\bar{a}(\mathbf{k})\bar{u}(x,\mathbf{k})\,, \tag{4.31a}$$

$$\bar{b}^{\dagger\dagger}(t,\mathbf{k})\bar{v}(\mathbf{x},\mathbf{k})=\bar{b}^\dagger(\mathbf{k})\bar{v}(x,\mathbf{k})\,, \tag{4.31b}$$

are consistent with (4.7b) and (4.8b).

We then define

$$\overline{\varphi}(x)^\mu = \int d^3k\, [\overline{a}(\mathbf{k})^\nu \overline{u}(x,\mathbf{k})^{\nu\mu} + \overline{b}^\dagger(\mathbf{k})^\nu \overline{v}(x,\mathbf{k})^{\nu\mu}]\,, \tag{4.32}$$

which satisfies

$$\overline{\varphi}(x)^\nu \lambda_\kappa(-\overleftarrow{\partial})^{\nu\mu} = 0\,. \tag{4.33}$$

The property of the divisor defined by (4.16) gives us

$$\overline{\varphi}(x)\{\overleftarrow{\partial}_t^2 + [\omega(i\overleftarrow{\nabla}) - i\kappa(i\overleftarrow{\nabla})A(i\overleftarrow{\nabla})]^2\} = 0\,. \tag{4.34}$$

E. The inner product of wave functions

Let us assume

$$\lambda_\kappa(\partial)^{\mu\nu} = \lambda^{(0)}(-i\nabla)^{\mu\nu} + i\lambda^{(1)}(-i\nabla)\delta^{\mu\nu} \times \partial_t + \lambda^{(2)}(-i\nabla)\delta^{\mu\nu}\partial_t^2\,. \tag{4.35}$$

The inner product of two wave functions, say $f(x)$ and $g(x)$, is defined by

$$(f\cdot g)_t^{\mu\nu} = \int d^3x \overline{f}(x)^{\mu\nu}\overleftrightarrow{\Gamma}(\partial,-\partial)^{\gamma\delta}g(x)^{\delta\nu}\,, \tag{4.36}$$

where we have introduced

$$\overleftrightarrow{\Gamma}(\partial,-\partial)^{\mu\nu} = [\lambda^{(1)}(-i\nabla) - i\lambda^{(2)}(-i\nabla)\overleftrightarrow{\partial}_t]\delta^{\mu\nu}\,, \tag{4.37}$$

with

$$\overleftrightarrow{\partial}_t = \partial_t - \overleftarrow{\partial}_t\,. \tag{4.38}$$

When f and g satisfy $\lambda_\kappa(\partial)f(x) = \lambda_\kappa(\partial)g(x) = 0$, we can show that $(f\cdot g)_t$ is independent of t as follows:

$$\begin{aligned}\frac{d}{dt}(f\cdot g)_t &= \int d^3x \overline{f}(x)[\lambda^{(1)}(-i\nabla)(\partial_t + \overleftarrow{\partial}_t) \\ &\quad - i\lambda^{(2)}(-i\nabla)(\partial_t + \overleftarrow{\partial}_t)(\partial_t - \overleftarrow{\partial}_t)]g(x) \\ &= -i\int d^3x \overline{f}(x)[\lambda_\kappa(\partial) - \lambda_\kappa(-\overleftarrow{\partial})]g(x) \\ &= 0\,,\end{aligned} \tag{4.39}$$

where we have performed integration by parts with respect to the space integration.

F. An orthonormalized complete set of solutions of the semifree field equation (4.14)

Being equipped with the above definition of inner product we now construct an orthonormalized complete set of solutions of the semifree field equation (4.14).

When we use $u(x,\mathbf{k})$ and $v(x,\mathbf{k})$ for $f(x)$ and $g(x)$ in (4.36), respectively, we have the following orthogonality theorem:

$$\int d^3x\, \overline{u}(x,\mathbf{k})\overleftrightarrow{\Gamma}_\kappa(\partial,-\partial)v(x,\mathbf{l}) = 0\,, \tag{4.40a}$$

$$\int d^3x\, \overline{v}(x,\mathbf{k})\overleftrightarrow{\Gamma}_\kappa(\partial,-\partial)u(x,\mathbf{l}) = 0\,, \tag{4.40b}$$

because of the time independence of the quantities proved by (4.39).

We choose $u(x,\mathbf{k})$ and $v(x,\mathbf{k})$ to satisfy the following orthonormalization condition:

$$\int d^3x\, \overline{u}(x,\mathbf{k})^{\mu\gamma}\overleftrightarrow{\Gamma}_\kappa(\partial,-\partial)^{\gamma\delta}u(x,\mathbf{l})^{\delta\nu} = \delta^{\mu\nu}\delta(\mathbf{k}-\mathbf{l})\,, \tag{4.41a}$$

$$\int d^3x\, \overline{v}(x,\mathbf{k})^{\mu\gamma}\overleftrightarrow{\Gamma}_\kappa(\partial,-\partial)^{\gamma\delta}v(x,\mathbf{l})^{\delta\nu} = -\rho\delta^{\mu\nu}\delta(\mathbf{k}-\mathbf{l})\,. \tag{4.41b}$$

In (4.41a) we chose the sign of $\lambda(\partial)$ (which determines the sign of $\overleftrightarrow{\Gamma}_\kappa$) in such a way that the left-hand side of (4.41a) is positive. Since this does not determine the sign of the quantity in (4.41b), we put the sign factor $\rho = \pm 1$ in the condition (4.41b).

Introducing

$$\Gamma[k_0,\mathbf{k}] = \lambda^{(1)}(\mathbf{k}) - 2k_0\lambda^{(2)}(\mathbf{k}) = \frac{\partial}{\partial k_0}\lambda[k_0,\mathbf{k}]\,, \tag{4.42a}$$

with

$$\lambda[k_0,\mathbf{k}] = \lambda^{(0)}(\mathbf{k}) + \lambda^{(1)}(\mathbf{k})k_0 - \lambda^{(2)}(\mathbf{k})k_0^2\,, \tag{4.42b}$$

for the case given by (4.35), we see that (4.41) give

$$\{\overline{u}(\mathbf{k})\Gamma[\omega(\mathbf{k}) - i\kappa(\mathbf{k})A(\mathbf{k}),\mathbf{k}]u(\mathbf{k})\}^{\mu\nu} = (2\pi)^{-3}\delta^{\mu\nu}\,, \tag{4.43a}$$

$$\{\overline{v}(\mathbf{k})\Gamma[-\omega(\mathbf{k}) + i\kappa(\mathbf{k})A(\mathbf{k}),-\mathbf{k}]v(\mathbf{k})\}^{\mu\nu} = -\rho(2\pi)^{-3}\delta^{\mu\nu}\,. \tag{4.43b}$$

G. The canonical sum rules

Let us define the following two functions:

$$\Delta_\kappa^\pm(x) = \mp i\int \frac{d^3k}{(2\pi)^3}\frac{1}{2[\omega(\mathbf{k}) - i\kappa(\mathbf{k})A(\mathbf{k})]} \times U_{\pm\kappa}(t,\mathbf{k})e^{\mp i[\omega(\mathbf{k})t - \mathbf{k}\cdot\mathbf{x}]}\,. \tag{4.44}$$

We then have

$$\Delta_\kappa^+(\mathbf{x},t) = -\Delta_\kappa^-(\mathbf{x},-t)\,, \tag{4.45}$$

$$\delta(t)\partial_t\Delta_\kappa^\pm(x) = -\tfrac{1}{2}\delta(\mathbf{x})\delta(t)\,, \tag{4.46}$$

$$\{\partial_t^2 + [\omega(-i\nabla) - i\kappa(-i\nabla)A(-i\nabla)]^2\}\Delta^\pm(x) = 0\,. \tag{4.47}$$

The last property implies that $d_\kappa(\partial)\Delta_\kappa^\pm(x)$ satisfies the equation

$$\lambda_\kappa(\partial)[d_\kappa(\partial)\Delta_\kappa^\pm(x)] = 0\,. \tag{4.48}$$

The following sum rules can be proved:

$$\int d^3k\, u(x,\mathbf{k})^{\mu\delta}\overline{u}(y,\mathbf{k})^{\delta\nu} = i\, d_\kappa(\partial)^{\mu\delta}\Delta_\kappa^+(x-y)^{\delta\nu}\,, \tag{4.49a}$$

$$\int d^3k\, v(x,\mathbf{k})^{\nu\delta}\overline{v}(y,\mathbf{k})^{\delta\nu} = -i\rho d_\kappa(\partial)^{\mu\delta}\Delta_\kappa^-(x-y)^{\delta\nu}\,, \tag{4.49b}$$

the detailed derivation of which is given in the Appendix. Since this sum rule is the basis of the equal-time canonical commutation relations, this is called the canonical sum rule.[8]

Introducing

$$\Delta_\kappa(x,t) = \Delta_\kappa^+(\mathbf{x},t) + \Delta_\kappa^-(\mathbf{x},t)\,, \tag{4.50}$$

we see from (4.45) and (4.46) that

$$\Delta_\kappa(\mathbf{x},t)\delta(t)=0\,, \tag{4.51}$$

$$\delta(t)\partial_t\Delta_\kappa(\mathbf{x},t)=-\delta(\mathbf{x})\delta(t)\,. \tag{4.52}$$

Since Δ_κ satisfies the linear homogeneous differential equation of the form (4.47) which is the second order in time derivative, we have

$$[(\partial_t)^{2n}\Delta_\kappa(\mathbf{x},t)]\delta(t)=0\,, \tag{4.53}$$

$$[(\partial_t)^{2n+1}\Delta_\kappa(\mathbf{x},t)]\delta(t)=0 \quad \text{for } \mathbf{x}\neq 0\,, \tag{4.54}$$

when n is an integer. Thus we have

$$[F(\partial)\Delta_\kappa(\mathbf{x},t)]\delta(t)=0 \quad \text{for } \mathbf{x}\neq 0\,, \tag{4.55}$$

where $F(\partial)$ stands for a sum of products of derivatives with *finite* powers. The combination $(\Delta_\kappa^+ - \Delta_\kappa^-)$ does not have this property.

H. The commutation relation and statistics

Using (4.23) and (4.32) with (4.49), we obtain

$$[\varphi(x)^\mu,\bar\varphi(y)^\nu]_\sigma = i\{d_\kappa(\partial)[\Delta_\kappa^+(x-y)+\sigma\rho\Delta_\kappa^-(x-y)]\}^{\mu\nu}\,. \tag{4.56}$$

We now require the causality condition which states that the operators of observables should commute with each other when they refer to different points in space at a common time. Then, it follows from (4.55) and (4.56) that

$$\sigma\rho=1\,. \tag{4.57}$$

In this way, ρ determines the statistics.

I. Projection of creation and annihilation operators

When we are given the semifree field $\varphi(x)$ of the form (4.23), we can project out the creation and annihilation operators by means of the formulas:

$$a(\mathbf{k})^\mu=\int d^3x\,\bar u(x,\mathbf{k})^{\mu\delta}\overleftrightarrow{\Gamma}(\partial,-\partial)^{\delta\nu}\varphi(x)^\nu\,, \tag{4.58a}$$

$$-\rho b^\dagger(\mathbf{k})^\mu=\int d^3x\,\bar v(x,\mathbf{k})^{\mu\delta}\overleftrightarrow{\Gamma}(\partial,-\partial)^{\delta\nu}\varphi(x)^\nu\,. \tag{4.58b}$$

In deriving (4.58), we have used (4.40) and (4.41).

J. The two-point Green's function

The internal line in the Feynman-type diagrams is the causal two-point Green's function $\Delta_c(x-y)$ defined by [see (4.18)]

$$\Delta_c(x-y)_{ij}^{\mu\nu}=d_\kappa(\partial)_{ij}^{\mu\delta}\Delta_c(x-y)^{\delta\nu} = -i\langle 0|T[\varphi(x)_i^\mu\bar\varphi(y)_j^\nu]|0\rangle\,. \tag{4.59}$$

Introducing the Fourier transform of $\Delta_c(x-y)$ with respect to the space variable by

$$\Delta_c(x-y)_{ij}^{\mu\nu}=\int\frac{d^3k}{(2\pi)^3}\Delta_c(t-s,\mathbf{k})_{ij}^{\mu\nu}e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}\,, \tag{4.60}$$

we have

$$\Delta_c(t-s,\mathbf{k})_{ij}^{\mu\nu}=G(t-s,\mathbf{k})^{\mu\delta}\sum_r[u_r(\mathbf{k})_i\bar u_r(\mathbf{k})_j]^{\delta\nu} + \sigma G(s-t,\mathbf{k})^{\mu\delta}\sum_r[v_r(\mathbf{k})_i\bar v_r(\mathbf{k})_j]^{\delta\nu}\,, \tag{4.61}$$

where

$$G(t-s,\mathbf{k})=G(t,s,\mathbf{k}) \tag{4.62}$$

with (2.23)–(2.27). Note the time independence of B, ω, and κ.

Fourier transforming (4.61) further with respect to time, we have

$$\Delta_c(k_0,\mathbf{k})_{ij}^{\mu\nu} = \int_{-\infty}^{\infty}dt\,\Delta_c(t,\mathbf{k})_{ij}^{\mu\nu}e^{ik_0t} = \left[\frac{1}{k_0-\omega(\mathbf{k})+i\kappa(\mathbf{k})A(\mathbf{k})}\sum_r u_r(\mathbf{k})_i\bar u_r(\mathbf{k})_j\right]^{\mu\nu} - \sigma\left[\frac{1}{k_0+\omega(\mathbf{k})-i\kappa(\mathbf{k})A(\mathbf{k})}\sum_r v_r(\mathbf{k})_i\bar v_r(\mathbf{k})_j\right]^{\mu\nu}. \tag{4.63}$$

In deriving (4.63), we have used the property

$$A(\mathbf{k})=B^{-1}(\mathbf{k})\tau_3B(\mathbf{k})\,. \tag{4.64}$$

Note that the existence of $i\kappa A$ in Δ_c means that the Feynman line is a dissipative wave even though no physical quantities dissipate in the stationary situation.

K. Some examples

A simple example of an equation of type II is given by

$$\lambda_\kappa(\partial)=-\partial_t^2-[\omega(-i\nabla)-i\kappa(-i\nabla)A(-i\nabla)]^2\,. \tag{4.65}$$

In this case,

$$d_\kappa(\partial)^{\mu\nu}=\delta^{\mu\nu}\,, \tag{4.66}$$

$$\overleftrightarrow{\Gamma}_\kappa(\partial,-\partial)^{\mu\nu}=\delta^{\mu\nu}i\overleftrightarrow{\partial}_t\,, \tag{4.67}$$

$$\rho=\sigma=1\,. \tag{4.68}$$

Thus (4.43) leads to

$$u(\mathbf{k})=\bar u(\mathbf{k})=v(\mathbf{k})=\bar v(\mathbf{k}) = (2\pi)^{-3/2}\{2[\omega(\mathbf{k})-i\kappa(\mathbf{k})A(\mathbf{k})]\}^{-1/2}\,. \tag{4.69}$$

From (4.49), the commutation relation becomes

$$[\varphi(x)^\mu,\bar\varphi(y)^\nu]_{+1}=i\Delta_\kappa(x-y)^{\mu\nu}\,, \tag{4.70}$$

where $\Delta_\kappa(x-y)$ is defined by (4.50). If we use the property (4.52), (4.70) reduces to

$$[\varphi(x)^\mu,\bar\pi(y)^\nu]_{+1}\delta(t-s)=i\delta(\mathbf{x}-\mathbf{y})\delta(t-s)\,, \tag{4.71}$$

where

$$\bar\pi(x)=\partial_t\bar\varphi(x)\,. \tag{4.72}$$

This shows that $\varphi(x)$ and $\bar\pi(x)$ are canonical conjugates of each other. Note that the canonical commutation relation is based on the canonical sum rule.

The propagator $\Delta_c(k_0,\mathbf{k})$ is given by

$$\Delta_c(k_0,\mathbf{k}) = B^{-1}(\mathbf{k})\frac{1}{k_0^2 - [\omega(\mathbf{k}) - i\kappa(\mathbf{k})\tau_3]^2}B(\mathbf{k})\,. \tag{4.73}$$

A more complicated example is provided by the semifree field equation of physical electrons in superconductors:

$$\lambda_\kappa(\partial)^{\mu\nu}_{ij} = i\,\partial_t\delta_{ij}\delta^{\mu\nu} - [\epsilon(-i\nabla)\tau_{3ij} - \Delta\tau_{1ij}] \times\frac{\omega(-i\nabla)\delta^{\mu\nu} - i\kappa(-i\nabla)A(-i\nabla)^{\mu\nu}}{\omega(-i\nabla)}, \tag{4.74}$$

with

$$\epsilon(-i\nabla) = (1/2m)[(-i\nabla)^2 - k_{\mathrm{F}}^2]\,, \tag{4.75}$$

$$\omega(-i\nabla)^2 = \epsilon(-i\nabla)^2 + \Delta^2\,, \tag{4.76}$$

where k_{F} is the Fermi momentum. It is easy to show that

$$d_\kappa(\partial)^{\mu\nu}_{ij} = i\,\partial_t\delta_{ij}\delta^{\mu\nu} + [\epsilon(-i\nabla)\tau_{3ij} - \Delta\tau_{1ij}] \times\frac{\omega(-i\nabla)\delta^{\mu\nu} - i\kappa(-i\nabla)A(-i\nabla)^{\mu\nu}}{\omega(-i\nabla)}. \tag{4.77}$$

From (4.37), we have

$$\overleftrightarrow{\Gamma}_\kappa(\partial,-\partial)^{\mu\nu}_{ij} = \delta_{ij}\delta^{\mu\nu}\,, \tag{4.78}$$

and

$$\rho = \sigma = -1\,. \tag{4.79}$$

Thus (4.43) leads to

$$u(\mathbf{k})^{\mu\nu}_i = \frac{1}{(2\pi)^{3/2}}\begin{pmatrix}\cos\theta(\mathbf{k})\\ -\sin\theta(\mathbf{k})\end{pmatrix}^{\delta^{\mu\nu}}_i, \tag{4.80a}$$

$$v(\mathbf{k})^{\mu\nu}_i = \frac{1}{(2\pi)^{3/2}}\begin{pmatrix}\sin\theta(\mathbf{k})\\ \cos\theta(\mathbf{k})\end{pmatrix}^{\delta^{\mu\nu}}_i, \tag{4.80b}$$

with

$$\cos\theta(\mathbf{k}) = \{[\omega(\mathbf{k}) + \epsilon(\mathbf{k})]/2\omega(\mathbf{k})\}^{1/2}\,, \tag{4.81a}$$

$$\sin\theta(\mathbf{k}) = \{[\omega(\mathbf{k}) - \epsilon(\mathbf{k})]/2\omega(\mathbf{k})\}^{1/2}\,. \tag{4.81b}$$

The commutation relation is given by (4.56) with (4.44) and (4.76). Furthermore, we have

$$[\varphi(x)^\mu_i,\bar\varphi(y)^\nu_j]_{-1}\delta(t-s) = \delta^{\mu\nu}\delta_{ij}\delta(\mathbf{x}-\mathbf{y})\delta(t-s)\,, \tag{4.82}$$

which indicates that $\varphi(x)$ and $\bar\varphi(x)$ form a pair of canonically conjugate fields. The propagator is given by (4.63) with (4.80).

A dissipative Dirac field satisfies the semifree field equation

$$\lambda_\kappa(\partial) = i\gamma^0\,\partial_t - (-i\gamma\cdot\nabla + m) \times\frac{\omega(-i\nabla) - i\kappa(-i\nabla)A(-i\nabla)}{\omega(-i\nabla)}, \tag{4.83}$$

with the Dirac γ matrices

$$\gamma^0 = \beta\,,\quad \gamma = \beta\alpha\,. \tag{4.84}$$

The $\omega(-i\nabla)$ is defined by

$$\omega(-i\nabla)^2 = (-i\nabla)^2 + m^2\,. \tag{4.85}$$

It is easy to show that

$$d_\kappa(\partial) = i\gamma^0\,\partial_t + (-i\gamma\cdot\nabla + m) \times\frac{\omega(-i\nabla) - i\kappa(-i\nabla)A(-i\nabla)}{\omega(-i\nabla)}, \tag{4.86}$$

so we have from (4.37)

$$\overleftrightarrow{\Gamma}_\kappa(\partial,-\partial)^{\mu\nu}_{ij} = (\gamma^0)_{ij}\delta^{\mu\nu}\quad (i,j = 1-4) \tag{4.87}$$

and

$$\rho = \sigma = -1\,. \tag{4.88}$$

From (4.43) we can construct the wave function as

$$u(\mathbf{k})^{\mu\nu}_i = u(\mathbf{k})_i\delta^{\mu\nu}\,, \tag{4.89a}$$

$$v(\mathbf{k})^{\mu\nu}_i = v(\mathbf{k})_i\delta^{\mu\nu}\,, \tag{4.89b}$$

where $u^{(r)}(\mathbf{k})_i$ and $v^{(r)}(\mathbf{k})_i$ $(r = 1,2)$ are four-component Dirac free spinors with the condition

$$\begin{aligned} &u^{(r)\dagger}(\mathbf{k})\gamma_0 u^{(s)}(\mathbf{k}) = \delta^{rs}/(2\pi)^3\,,\\ &v^{(r)\dagger}(\mathbf{k})\gamma_0 v^{(s)}(\mathbf{k}) = \delta^{rs}/(2\pi)^3\,,\\ &u^{(r)\dagger}(\mathbf{k})\gamma_0 v^{(s)}(\mathbf{k}) = v^{(r)}(\mathbf{k})^\dagger\gamma_0 u^{(s)}(\mathbf{k}) = 0\,. \end{aligned} \tag{4.90}$$

Again the commutation relation is given by (4.56) with (4.44) and (4.86). It leads to

$$[\varphi(x)^\mu_i,\bar\varphi(y)^\nu_j]_{-1}\delta(t-s) = \delta^{\mu\nu}(\gamma_0)_{ij}\delta(\mathbf{x}-\mathbf{y})\delta(t-s)\,, \tag{4.91}$$

which indicates φ and $\bar\varphi\gamma_0$ are canonical conjugates of each other.

In this section we used the particular choice, $\alpha = \frac{1}{2}$, in constructing the canonical formalism of the semifree fields. The construction is tremendously simplified by the relation $\tau_2 A^T\tau_2 = -A$, which holds if and only if $\alpha = \frac{1}{2}$. The construction of the semifree field with arbitrary α requires a more complex consideration. This will be presented elsewhere.

V. SELF-CONSISTENT EQUATIONS OF ϕ^3 MODEL

In this section we present an example of ϕ^3 self-interacting real scalar in order to show how a dissipative effect can be created spontaneously in such an isolated system in TFD. (This system is assumed to be stationary below.)

In a recent paper[7] we have already seen spontaneous creation of dissipation in a reservoir model, a simple solvable one, which is a system of a single harmonic oscillator bilinearly coupled to a reservoir consisting of an infinite number N of harmonic oscillators. At the limit of $N\to\infty$ with fixed $\bar g^2$,

$$\bar g^2 = Ng^2\,, \tag{5.1}$$

g being a coupling constant between the system and reservoir, our self-consistent renormalization scheme leads us to a solution in which κ is nonvanishing and the number density of the system is determined to be at the temperature of the reservoir. The infiniteness of degrees of freedom in the reservoir, allowing the system to dissipate in it, is an essential ingredient for this. This analysis also tells us that the dissipation effect shows up as a result of the communication among the tilde and nontilde fields and, consequently, it can really be called thermal dissipation.

When we consider a nonlinear self-interacting case which is being dealt with in what follows, we should take account of the self-energy diagrams due to the self-interaction in the self-consistent equations. Although the system is no longer coupled to any reservoir we still have an infinite number of communication channels among the tilde and nontilde field operators due to transitions from one state to another by self-interaction. In this case as well as general nonlinearly interacting cases we may expect the thermal instability, i.e., that nontilde particles decay into multiparticles including tilde particles through many channels, because tilde particles have negative energies. So the thermal dissipation inevitably appears in nonlinearly interacting cases.

The formulation of the semifree field of type II in a stationary system has been given in Sec. IV with the choice of $\alpha = \frac{1}{2}$. We start from the model Lagrangian density for a real scalar field,

$$\mathscr{L} = \tfrac{1}{2}[\dot{\varphi}_0^2 - \varphi_0\omega_0^2(-i\nabla)\varphi_0] - (g_0/3!)\varphi_0^3, \quad (5.2)$$

where ω_0, g_0, and φ_0 are bare quantities. Deriving H from the canonical formula and constructing its tilde conjugate $\widetilde{H}$, we have our basic total Hamiltonian $\widehat{H} = H - \widetilde{H}$. The semifree (unperturbed renormalized) Hamiltonian $\widehat{H}^0$, which specifies the state vector space for realization of the field operator, should have the following form:

$$\widehat{H}^0 = \int d^3\mathbf{x}\, \frac{1}{2}[\dot{\overline{\varphi}}^\mu \dot{\varphi}^\mu + \overline{\varphi}^\mu\{(\omega - i\kappa A)^2\}^{\mu\nu}\varphi^\nu], \quad (5.3)$$

where φ is the semifree (unperturbed renormalized) field [see (4.23)] and ω is a renormalized energy. Then the interaction Hamiltonian $\widehat{H}_I$ is unambiguously given by

$$\widehat{H}_I = \widehat{H} - \widehat{H}^0 = \widehat{H}_{\text{int}} + \widehat{H}_c, \quad (5.4a)$$

$$\widehat{H}_{\text{int}} = \frac{g}{3!}\int d\mathbf{x}[\{\varphi^1\}^3 - \{\varphi^2\}^3], \quad (5.4b)$$

$$\widehat{H}_c = \frac{1}{2}\int d\mathbf{x}\, \overline{\varphi}^\mu[-\delta\omega^2 + 2i\kappa\omega A + \kappa^2]^{\mu\nu}\varphi^\nu, \quad (5.4c)$$

where g is a renormalized coupling constant and an energy counterterm $\delta\omega^2$ is defined by

$$\omega_0^2 = \omega^2 - \delta\omega^2. \quad (5.4d)$$

In writing Eq. (5.4c), we used the relation

$$A^2 = 1, \quad (5.4e)$$

and suppressed the counterterms of wave function renormalization and the coupling constant renormalization since we are interested only in the two-point Green's functions in the one-loop approximation.

In order to apply the self-consistent renormalization scheme to this system, we now calculate the connected full propagator $\Delta_{c,\text{full}}$,

$$\Delta_{c,\text{full}}(x-y)^{\mu\nu} = -i\langle 0|T\left[\varphi(x)^\mu\overline{\varphi}(y)^\nu \times \exp\left\{-i\int dt\,\widehat{H}_I(t)\right\}\right]|0\rangle_{\text{conn}} \quad (5.5)$$

where the suffix conn means a connected part of the diagrams. From graphical considerations just as in the ordinary quantum field theory, $\Delta_{c,\text{full}}^{-1}$ can be expressed by the proper self-energy Σ as

$$\Delta_{c,\text{full}}^{-1}(k)^{\mu\nu} = \Delta_c^{-1}(k)^{\mu\nu} - \Sigma(k)^{\mu\nu}, \quad (5.6)$$

the free propagator Δ_c being given in (4.72) and $k = (k_0, \mathbf{k})$. At the level of one-loop approximation, $\Sigma(k)$ is a sum of the contribution from a one-loop diagram $\Sigma_l(k)$ and that from counterterms,

$$\Sigma(k)^{\mu\nu} = \Sigma_l(k)^{\mu\nu} + \{-\delta\omega^2 + \kappa^2 + 2i\kappa\omega A\}(\mathbf{k})^{\mu\nu}, \quad (5.7a)$$

$$\Sigma_l(k)^{\mu\nu} = \frac{g^2}{2}\int\frac{d^4q}{(2\pi)^4} \times \begin{bmatrix}\Delta(q_+)^{11}\Delta(q_-)^{11} & -\Delta(q_+)^{12}\Delta(q_-)^{12}\\ \Delta(q_+)^{21}\Delta(q_-)^{21} & -\Delta(q_+)^{22}\Delta(q_-)^{22}\end{bmatrix},$$

$$q_\pm = q \pm k/2. \quad (5.7b)$$

For a real value of k_0, the integrations over q_0 can be performed in (5.7b), and Σ_l has the matrix form of

$$\Sigma_l(k) = \begin{bmatrix}L_1 - iL_2 & iL_3\\ -iL_3 & L_1 + iL_2\end{bmatrix}, \quad (5.8a)$$

whose elements are further expressed as

$$L_1 = -\frac{g^2}{2}\int\frac{d^3\mathbf{q}}{(2\pi)^3} \times\{(1+2n_+)K''_{(+)} + (1+2n_-)K''_{(-)}\}, \quad (5.8b)$$

$$L_2 = -\frac{g^2}{2}\int\frac{d^3\mathbf{q}}{(2\pi)^3} \times\{K'_{(+)} + (1+2n_+)(1+2n_-)K'_{(-)}\}, \quad (5.8c)$$

$$L_3 = -\frac{g^2}{2}\int\frac{d^3\mathbf{q}}{(2\pi)^3} \times\{4\sqrt{n_+(1+n_+)n_-(1+n_-)}\,K'_{(-)}\}, \quad (5.8d)$$

with the notation of $n_\pm = n(\mathbf{q}\pm\mathbf{k}/2)$. The real valued functions $K'_{(\pm)}$ and $K''_{(\pm)}$ are defined through the following relations:

$$K'_{(\pm)} + iK''_{(\pm)} \equiv \tfrac{1}{2}(I_1 \pm I_2), \quad (5.9a)$$

where

$$I_1 = \frac{-i(\Omega_+ + \Omega_-)}{2\Omega_+\Omega_-[k_0^2 - (\Omega_+ + \Omega_-)^2]}, \quad (5.9b)$$

$$I_2 = \frac{i(\Omega_+ - \Omega_-^*)}{2\Omega_+\Omega_-^*[k_0^2 - (\Omega_+ - \Omega_-)^2]}, \quad (5.9c)$$

$$\Omega_\pm = \omega_\pm - i\kappa_\pm,$$

with $\omega_\pm = \omega(\mathbf{q}\pm\mathbf{k}/2)$ and $\kappa_\pm = \kappa(\mathbf{q}\pm\mathbf{k}/2)$.

At this stage, we require the renormalization condition that the total self-energy should vanish on the energy shell, $k_0 = \omega(\mathbf{k})$,

$$\Sigma(k_0 = \omega(\mathbf{k})) = 0 \quad (5.10a)$$

or

$$-\Sigma_l(k_0 = \omega(\mathbf{k})) = \{-\delta\omega^2 + \kappa^2 + 2i\kappa\omega A\}(\mathbf{k})^{\mu\nu}. \quad (5.10b)$$

The last equation is the 2×2 matrix self-consistent equation. In deriving (5.10), the "on shell" is defined by the real part of the pole of propagator. It is remarkable that this matrix equation brings us only three independent real equations because the two off-diagonal elements give the same equation,

$$4\kappa\omega\sqrt{n(l+n)} = L_3(k_0=\omega,\mathbf{k}) . \quad (5.11a)$$

The real and imaginary part of the two diagonal elements of (5.10b) imply

$$\delta\omega^2 - \kappa^2 = L_1(k_0=\omega,\mathbf{k}) \quad (5.11b)$$

and

$$2n\omega(1+2n) = L_2(k_0=\omega,\mathbf{k}) , \quad (5.11c)$$

respectively. In Eqs. (5.11), we used the notations $n=n(\mathbf{k})$, $\omega=\omega(\mathbf{k})$, $\kappa=\kappa(\mathbf{k})$, and $\delta\omega=\delta\omega(\mathbf{k})$.

The set of three equations (5.11) are the self-consistent equations for three unknown functions $\omega(\mathbf{k})$, $n(\mathbf{k})$, and $\kappa(\mathbf{k})$. Their solution is expressed in terms of g_0 and m_0, where

$$\omega_0^2 = \mathbf{k}^2 + m_0^2 . \quad (5.12)$$

We have been unable so far to obtain analytic solutions except for $\kappa(\mathbf{k})=0$, because the forms of the functions $\omega(\mathbf{k})$, $\kappa(\mathbf{k})$, and $n(\mathbf{k})$ are not given but should be determined self-consistently. However, we expect that there are many solutions with $\kappa\neq 0$, each having different $\omega(\mathbf{k})$, $\kappa(\mathbf{k})$, and $n(\mathbf{k})$. The $n(\mathbf{k})$ thus obtained is determined by dynamics which includes the interaction with the thermal background fields (i.e., the tilde quanta).

The origin of $\kappa=0$ solution in our calculation can be understood in the following way. Suppose that we start our perturbation calculation with the free fields without dissipation and consider the self-energy at one-loop level. Then the decay of a nontilde particle into two nontilde particles is obviously forbidden. At first glance, we might feel that the decay of a nontilde particle into another nontilde particle and a tilde particle might be possible because the tilde particle we consider has a negative energy. The energy and momentum conservation laws read as

$$\omega(\mathbf{P}) = \omega(\mathbf{k}_1) - \omega(\mathbf{k}_2) , \quad (5.13a)$$

$$\mathbf{P} = \mathbf{k}_1 - \mathbf{k}_2 , \quad (5.13b)$$

with

$$\omega(\mathbf{P}) = \sqrt{\mathbf{P}^2 + m^2} . \quad (5.13c)$$

However, further inspection of (5.13) shows that there is no solution of (5.13). This means that the $\kappa=0$ solution in the self-consistent equation can be interpreted as the result of the *lowest-order* perturbation expressed in terms of the usual free field without dissipation. When we consider higher orders, a nontilde particle is allowed to decay into many particle states with at least one tilde particle and, consequently, *$\kappa=0$ is no longer a solution of the self-consistent equation.*

VI. A SHORT SUMMARY

It was shown that TFD offers a systematic and unified treatment of any thermal situations including nonequilibrium situations. Given a dynamics, i.e., a Hamiltonian, H characterizing a system under study, one associates with it the total Hamiltonian $\hat{H}$ as $\hat{H} = H - \tilde{H}$. Each thermal situation corresponds to one of various realizations of $\hat{H}$ and is characterized by thermal state conditions at an initial time. With such initial thermal state conditions we proceed to make an interaction picture associated with it and follow a self-consistent renormalization program to determine the time evolution of n as well as ω and κ. In other words, all the quantities n, ω, and κ are determined dynamically and self-consistently. Note that here the dynamics include the communication between the nontilde particles and the thermal background fields (the tilde fields).

In the course of practical calculations we can exploit all the techniques developed in the usual field theory, because TFD is formulated as an operator field theory. Thus the use of TFD will be useful in describing the nonequilibrium system with an infinite number of degrees of freedom like in cases of quantum field systems.

Since TFD is equivalent to the other methods such as the density matrix formalism[13] and the path-ordering method,[12] we expect that a formulation of self-consistent renormalization for the dissipation coefficient may be needed also in other methods.

One particular flexibility in TFD can be found in the arbitrariness of the parameter α discussed in Sec. III. The specific choice of α sometimes simplifies a problem in a way similar to the choice of gauge in gauge theories.

We pointed out in Sec. I how the thermal instability in TFD is closely related to the lower unboundedness of $\hat{H}$. It indicates that the dynamical map in terms of the asymptotic free fields is inadequate in TFD. Therefore, we have formulated the semifree fields in $[\alpha=\frac{1}{2}]$ representation for unperturbed ones, taking account of such a thermal instability from the beginning. There it is remarkable that the semifree fields has a canonical formalism. This enables us to follow a self-consistent renormalization, since the renormalization transformation is a kind of canonical transformation.

In this paper we give only the semifree fields of type II in the $[\alpha=\frac{1}{2}]$ representation. The semifree field formulation in any $[\alpha]$ representation is possible, which is explicitly shown, e.g., in Ref. 8. We think that practical calculations are performed most elegantly in the present formulation of the $[\alpha=\frac{1}{2}]$ representation because of its symmetric nature.

The infinite number of degrees of freedom combined with the lower unboundedness of $\hat{H}$ plays a central role in the present formulation of TFD and gives us a much richer structure of theory than the quantum field theory without thermal degrees of freedom. First, the fact that all thermal situations should be covered by a single Hamiltonian $\hat{H}$ is justified by the existence of inequivalent representations inherent to quantum theory with infinite degrees of freedom. Second, the nonvanishing κ really appears as a result of infinite decay channels due to the negative energy of tilde particles. We call it spontaneous creation of dissipation, since its mechanism is analogous to the spontaneous breakdown of symmetry in quantum field theory. Third, although the complex eigenvalues of $\hat{H}^0$ seem to contradict the Hermiticity of $\hat{H}$, this controversy may be explained by the spontaneous breakdown of symmetry generated by the generator H in which H has no eigenstates and eigenvalues in the realization Fock space.

The theory developed in this paper still has many problems to be studied in the future. Above all, its application to the explicitly time-dependent case is of particular interest. There the physical content of the theory will manifest itself most clearly. After its accomplishment, the theory is applicable to the problem of the universe as well as various nonequilibrium problems in solid state physics. To this end, we need the formalisms of both canonical semifree fields and renormalization in time-dependent cases, on which we are preparing another paper.

APPENDIX: DERIVATION OF (4.49)

Since $d_\kappa(\partial)\Delta_\kappa^\pm(x-y)$ are solutions of field equation (4.14) [see (4.48)], they can be expanded in terms of the orthornormalized solutions $u(x,\mathbf{k})$ and $v(x,\mathbf{k})$ as

$$d_\kappa(\partial)\Delta_\kappa^+(x-y)=\int d^3k\, u(x,\mathbf{k})C^+(y,\mathbf{k})\,, \quad \text{(A1a)}$$

$$d_\kappa(\partial)\Delta_\kappa^-(x-y)=\int d^3k\, v(x,\mathbf{k})C^-(y,\mathbf{k})\,. \quad \text{(A1b)}$$

When we consider the orthonormalization condition (4.41), the matrices $C^\pm(y,\mathbf{k})$ can be determined as follows:

$$C^+(y,\mathbf{k})=\int d^3k\, \bar{u}(x,\mathbf{k})\overleftrightarrow{\Gamma}(\partial,-\partial)d(\partial)\Delta_\kappa^+(x-y)\,, \quad \text{(A2a)}$$

$$C^-(y,\mathbf{k})=\int d^3k\, \bar{v}(x,\mathbf{k})\overleftrightarrow{\Gamma}(\partial,-\partial)d(\partial)\Delta_\kappa^-(x-y)\,. \quad \text{(A2b)}$$

Substituting (4.44) into (A2a), we obtain

$$\begin{aligned}C^+(y,\mathbf{k})=&-i\frac{1}{2[\omega(\mathbf{k})-i\kappa(\mathbf{k})A(\mathbf{k})]}\\&\times\bar{u}(y,\mathbf{k})\Gamma[\omega(\mathbf{k})-i\kappa(\mathbf{k})A(\mathbf{k}),\mathbf{k}]\\&\times d\,[\omega(\mathbf{k})-i\kappa(\mathbf{k})A(\mathbf{k}),\mathbf{k}]\,. \quad \text{(A3)}\end{aligned}$$

On the other hand, we see from (4.42a) that

$$\Gamma[k_0,\mathbf{k}]d\,[k_0,\mathbf{k}]+\lambda\,[k_0,\mathbf{k}]\frac{\partial}{\partial k_0}d\,[k_0,\mathbf{k}]=2k_0\,, \quad \text{(A4a)}$$

where $d[k_0,\mathbf{k}]$ is defined by

$$\begin{aligned}\lambda[k_0,\mathbf{k}]d[k_0,\mathbf{k}]&=d[k_0,\mathbf{k}]\lambda[k_0,\mathbf{k}]\\&=k_0^2-[\omega(\mathbf{k})-i\kappa(\mathbf{k})A(\mathbf{k})]^2\,. \quad \text{(A4b)}\end{aligned}$$

This leads to

$$\begin{aligned}&\bar{u}(y,\mathbf{k})\Gamma[\omega(\mathbf{k})-i\kappa(\mathbf{k})A(\mathbf{k}),\mathbf{k}]d\,[\omega(\mathbf{k})-i\kappa(\mathbf{k})A(\mathbf{k}),\mathbf{k}]\\&\quad=2[\omega(\mathbf{k})-i\kappa(\mathbf{k})A(\mathbf{k})]\bar{u}(y,\mathbf{k})\,, \quad \text{(A5)}\end{aligned}$$

Thus (A3) gives

$$C^+(y,\mathbf{k})=-i\bar{u}(y,\mathbf{k})\,. \quad \text{(A6a)}$$

Similarly we have

$$C^-(y,\mathbf{k})=i\rho\bar{v}(y,\mathbf{k})\,. \quad \text{(A6b)}$$

Substituting (A6) into (A1), we obtain the sum rule (4.49).

[1] T. Arimitsu and H. Umezawa, Prog. Theor. Phys. **74**, 429 (1985); **77**, 32 (1987).
[2] T. Arimitsu and H. Umezawa, Prog. Theor. Phys. **77**, 53 (1987).
[3] T. Arimitsu, J. Pradko, and H. Umezawa, Physica A **135**, 487 (1986); H. Umezawa and T. Arimitsu, *Path Integrals from meV to MeV*, edited by M. C. Gutzwiller, A. Inomata, J. R. Klauder, and L. Streit (World Scientific, Singapore, 1986), p. 217; Prog. Theor. Phys. Suppl. **86**, 243 (1986); T. Arimitsu, Y. Sudo, and H. Umezawa, "Dynamical rearrangement of the thermal vacuum in TFD," Physica A, in press; H. Umezawa, *Symmetries in Science II*, edited by B. Gruber and R. Lenczewski (Plenum, New York, 1986), p. 517.
[4] T. Arimitsu and H. Umezawa, J. Phys. Soc. Jpn. **55**, 1475 (1986).
[5] T. Arimitsu, M. Guida, and H. Umezawa, "Dissipative quantum field theory—thermo field dynamics—," Physica A, in press.
[6] T. Arimitsu and H. Umezawa, *Advances on Phase Transitions and Disordered Phenomena*, edited by G. Busiello *et al.* (World Scientific, Singapore, 1987), p. 483; T. Arimitsu, M. Guida, and H. Umezawa, Europhys. Lett. **3**, 277 (1987); H. Umezawa and T. Arimitsu, *Proceedings of the 2nd International Symposium on the Foundation of Quantum Mechanics—In the Light of New Technology*, edited by M. Namiki *et al.* (Physical Society of Japan, Tokyo, 1987), p. 79.
[7] T. Arimitsu, H. Umezawa, Y. Yamanaka, and N. J. Papastamatiou, "Spontaneous creation of dissipation in thermo field dynamics and its examples," Physica A, in press.
[8] H. Umezawa, Y. Yamanaka, I. Hardman, and T. Arimitsu, *Wandering in the Fields*, edited by K. Kawarabayashi and A. Ukawa (World Scientific, Singapore, 1987).
[9] H. Umezawa, H. Matsumoto, and M. Tachiki, *Thermo Field Dynamics and Condensed States* (North-Holland, Amsterdam, 1982); see also N. P. Landsman and Ch. G. van Weert, Phys. Rep. **145**, 141 (1987).
[10] H. Umezawa, Prog. Theor. Phys. Suppl. **80**, 26 (1984); H. Matsumoto, *Progress in Quantum Field Theory*, edited by H. Ezawa and S. Kamefuchi (North-Holland, Amsterdam, 1985), p. 171.
[11] N. Narnhofer, M. Requardt, and W. Thirring, Commun. Math. Phys. **92**, 247 (1983).
[12] J. Schwinger, J. Math. Phys. **2**, 407 (1961); L. V. Keldysh, Soviet Phys. JETP **20**, 1018 (1965); R. A. Craig, J. Math. Phys. **9**, 605 (1968); R. Mills, *Propagators for Many-Particle Systems* (Gordon and Breach, New York, 1969).
[13] For example, F. Haake, *Springer Tracts in Modern Physics*, edited by G. Höhler (Springer, Berlin, 1973), Vol. 66, p. 98; see also, T. Arimitsu, J. Phys. Soc. Jpn. **51**, 1720 (1982).

Classical nonlinear σ models on Grassmann manifolds of compact or noncompact type

J-P. Antoine and B. Piette[a)]
Institut de Physique Théorique, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

(Received 17 March 1987; accepted for publication 15 July 1987)

Grassmannian σ models are reexamined in the light of a new geometrical result. Namely, the Cartan immersion of a Riemannian symmetric space G/H into its isometry group G is not always a diffeomorphism *onto* the set M_σ introduced by Eichenherr and Forger [Nucl. Phys. B **164**, 528 (1980)]. This is the case, in particular, for Grassmann manifolds of all types: their set M_σ is shown to be disconnected and its structure is completely analyzed. As a consequence, a new constraint must be introduced in the Bäcklund transformation (BT) method proposed by Harnad, Saint-Aubin, and Shnider [Commun. Math. Phys. **92**, 329 (1984)]. It is shown, however, to always be satisfied for Grassmann manifolds of compact type (i.e., G compact), but the problem remains open in most other cases. On the other hand, the BT method is extended to the Euclidean regime.

I. INTRODUCTION

Classical two-dimensional nonlinear σ models have been around for many years. They still remain popular, because they are nice examples of integrable systems, with many interesting properties: dual symmetry,[1-3] infinitely many conservation laws[4] with corresponding generators,[5,6] striking resemblance to four-dimensional Yang–Mills systems, etc.

The original impetus was the discovery by Pohlmeyer[1] that the S^n model possesses the so-called dual symmetry; the model describes a two-dimensional massless field constrained to live on a sphere S^n, but otherwise free. The obvious generalization is to replace the sphere $S^n = SO(n+1)/SO(n)$ by an arbitrary homogeneous space $M = G/H$, where G is a Lie group and H a closed subgroup of G. However, and this is the key point of the theory, Eichenherr and Forger[2,3] have shown that dual symmetry holds iff G/H is a Riemannian symmetric space (RSS). For convenience, we repeat here the definition of a (irreducible) symmetric space[7]: G is a classical Lie group, σ an involutive automorphism of G, and H a closed subgroup of G such that

$$(G_\sigma)_0 \subset H \subset G_\sigma, \tag{1.1}$$

where G_σ is the set of fixed points of σ and $(G_\sigma)_0$ its identity component. The relation (1.1) implies the canonical, σ_*-invariant, decomposition of $\mathfrak{g}$, the Lie algebra of G,

$$\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{m}, \tag{1.2}$$

where $\mathfrak{h}$ is the Lie algebra of H and one has

$$[\mathfrak{h},\mathfrak{h}] \subset \mathfrak{h}, \quad [\mathfrak{h},\mathfrak{m}] \subset \mathfrak{m}, \quad [\mathfrak{m},\mathfrak{m}] \subset \mathfrak{h}, \tag{1.3}$$

i.e., $\mathfrak{g}$ is reductive. The symmetric space G/H is Riemannian if its canonical G-invariant metric, induced by the Killing form of $\mathfrak{g}$, is positive definite, pseudo-Riemannian if the metric is indefinite.

Then, given a symmetric space $M = G/H$, the nonlinear σ model on M is the "free" field theory of a massless field ϕ: $\mathbb{R}^2 \to M$, where $\mathbb{R}^2$ may have either a Minkowskian or a Euclidean metric (we reproduce here the bare essentials only, referring the reader to the many existing reviews for further details, e.g., Ref. 4). The σ models come in two types.

(i) *Principal models*, in the case where M is itself a Lie group, $M \cong G = G \times G/G_{\rm diag}$. With the field denoted g: $\mathbb{R}^2 \to G$, the model is defined by the action

$$S = \frac{1}{2}\int_{\mathbb{R}^2} d^2x \, {\rm Tr}(g^{-1}\partial_\mu g)(g^{-1}\partial^\mu g), \quad \mu = 0,1, \tag{1.4}$$

and the equation of motion reads

$$\partial_\mu (g^{-1}\partial^\mu g) = 0. \tag{1.5}$$

(ii) *Nonprincipal models*, in the other cases. Then the field may be taken as $g(x) \in G$ as before, but subject to some additional constraints, and similarly for the equation of motion (1.5). Alternatively, one may also introduce the covariant derivative $D^\mu g$ as the horizontal part of $\partial^\mu g$ in the bundle $G \overset{H}{\to} G/H$, and transform Eq. (1.5) accordingly (for Grassmannian σ models, a more convenient parametrization will be used in Sec. V).

The crucial fact linking the two types of models is the so-called *Cartan immersion*[3,8-10]:

$$i_\sigma: G/H \to G,$$
$$i_\sigma(gH) = \sigma(g)g^{-1}, \tag{1.6}$$

by which the symmetric space G/H is mapped into its isometry group G; moreover the image $i_\sigma(G/H)$ is a closed totally geodesic submanifold of G. Then, as shown by Eichenherr and Forger,[3] the solutions of the G/H model are exactly those solutions of the principal G model that belong to the image $i_\sigma(G/H)$. The main tool in their analysis is the following set:

$$M_\sigma = \{ g \in G \,|\, \sigma(g)g = 1 \}, \tag{1.7}$$

which satisfies the obvious inclusions

$$i_\sigma(G/H) \subset M_\sigma \subset G. \tag{1.8}$$

However, contrary to the claim of Ref. 3 G/H is in general *not* globally diffeomorphic to M_σ. In all classical cases,[7] G/H is a connected manifold, but in many of them, M_σ is not

[a)] Chercheur IISN (Belgium).

 0022-2488/87/112753-10$02.50

connected, and not even a manifold! Actually each connected component of M_σ *is* a submanifold of G, but different components may have different dimensions. With a slight abuse of language, we may still call M_σ the EF *submanifold*. Then the result of Eichenherr and Forger[3] if that the Cartan immersion i_σ is a *local* diffeomorphism onto the identity component $(M_\sigma)_0$ of the EF submanifold M_σ.

The argument of Ref. 3 rests on the fact that the map Ψ: $G \to G$ given by $\Psi(Q) = \sigma(Q)Q$ has constant rank. However this is incorrect. Indeed the kernel of the tangent map Ψ_* at Q is given by $Q\,\mathfrak{k}_Q$, where $\mathfrak{k}_Q$ is the set of $X \in \mathfrak{m}$ that verify the equation

$$\sigma_*(X) + \mathrm{Ad}(Q)X = 0. \tag{1.9}$$

For $Q = \mathbb{1}$, (1.9) means $X \in \mathfrak{m}$ and thus rank $\Psi|_1 = \dim \mathfrak{h}$. But this is not true for *all* $Q \in G$. Indeed, for *each* classical symmetric space, as listed, e.g., in Refs. 7 and 11, we can find $Q_0 \in G$ such that rank $\Psi|_{Q_0} \neq \dim \mathfrak{h}$. More precisely, we have the following.

(i) If the automorphism σ is a simple conjugation, $\sigma(g) = LgL^{-1}$, then for $Q_0 = L$, Eq. (1.9) has the only solution $X = 0$ and rank $\Psi|_L = \dim \mathfrak{g}$. This is true for all Grassmannian spaces, i.e., the series AIII, BDI, and CII, and also for the *compact* series DIII and CI. In each case, L indeed belongs to G, except maybe for the fact that $\det L = -1$ for some dimensions; in that case, it is easy to find a modified L' that will do the job and yield rank $\Psi|_{L'} \neq \dim \mathfrak{h}$.

(ii) In all other cases, one may choose $Q_0 = J$, or more generally, $Q_0 = J'$, where, for appropriate values of m and k,

$$J \equiv J_m = \begin{bmatrix} 0 & \mathbb{1}_m \\ -\mathbb{1}_m & 0 \end{bmatrix}, \quad J' = \begin{bmatrix} \mathbb{1}_k & 0 \\ 0 & J_m \end{bmatrix}. \tag{1.10}$$

The conclusion is that rank Ψ is never constant, and the constant rank theorem does not apply! We may also notice that Q_0 belongs to M_σ in the Grassmannian case, but not for the other ones; for instance $\sigma(J)J = -\mathbb{1}$ in all cases.

In addition, the Cartan immersion is not always one-to-one. Indeed for $H \neq G_\sigma$ in Eq. (1.1), G/H is not simply connected, but a finite covering of G/G_σ, and then $(M_\sigma)_0$ is globally diffeomorphic to G/G_σ rather than to G/H itself. This point is discussed in the Erratum to Ref. 3, and will not occupy us any longer. We will encounter an example in Sec. IV below, namely the case of oriented real Grassmann manifolds.

To summarize, the reasoning of Ref. 3 effectively shows that *each* connected component of M_σ is a closed totally geodesic submanifold of G and the Cartan immersion i_σ is a diffeomorphism of G/G_σ onto the identity component $(M_\sigma)_0$.

Unfortunately, the result originally stated by Eichenherr and Forger is the cornerstone of most of the later developments of the theory of σ models, in particular the method of construction of multisoliton solutions introduced by Saint-Aubin[9,10] and Harnad, Saint-Aubin, and Shnider,[11] and extended by the same authors to general integrable systems using the so-called soliton correlation matrix.[12] Thus it is urgent to reexamine the validity of those methods in the light of the geometrical complications discussed above.

The aim of the present paper is double.

(i) First we want to clarify the geometrical setup. A complete analysis seems difficult for a general symmetric space, but we will perform it when M is a *Grassmann manifold*, complex, real, or quaternionic, of compact or noncompact type (in this second case, the space may be pseudo-Riemannian). This is the content of Secs. II–IV.

(ii) Then we reexamine in Sec. VI the Bäcklund transformation (BT) method of Harnad, Saint-Aubin, and Shnider[11,12] (those papers will be denoted HSS in the sequel) in the light of those results. The outcome is that the method is valid for Grassmannian σ models of compact type, and also for the σ models on the Riemannian spaces SU(n)/SO(n) and SU($2n$)/Sp(n), for which M_σ *is* connected. In all other cases, the question remains open.

In addition to those crucial points, we present in Sec. V a unified formulation of all types of Grassmannian models, compact or not, in terms of projection valued fields, as originally introduced by Zakharov and Mikhailov,[13] Corrigan *et al.*,[14] and Dubois-Violette and Georgelin.[15] In Sec. VI we also establish the validity of the HSS method for Euclidean models (only the Minkowskian case is considered in Refs. 11 and 12); this was taken for granted by Sasaki,[16] but the result is far from obvious. Incidentally the same paper contains some remarks about noncompact Grassmannian σ models, but the discussion is very sketchy and offers no proof whatsoever.

Of course the most popular approach for solving Euclidean σ models is the "holomorphic" method introduced by Borchers and Garber[17] and developed systematically by Din and Zakrzewski (see Refs. 18 and 19 for a review), and by Sasaki.[20] It turns out that this method also may be extended to noncompact models and their supersymmetric extensions. We will report on this elsewhere.

II. GEOMETRY OF THE EF SUBMANIFOLD: COMPACT GRASSMANNIANS

We begin with the simplest case, the complex Grassmann manifold of compact type,

$$G_{pq}(\mathbb{C}) = \mathrm{U}(p+q)/\mathrm{U}(p)\times\mathrm{U}(q) \tag{2.1}$$

$$= \mathrm{SU}(p+q)/S(\mathrm{U}(p)\times\mathrm{U}(q)). \tag{2.2}$$

In the sequel we will consider mainly the realization (2.1) of G_{pq}, despite the fact that $\mathrm{U}(p+q)$ does not act effectively on it (see the discussion of Forger[4] on this point). At the end of the section we will indicate the changes required by (2.2).

The involution σ_{pq} of G_{pq} is given[7,11] by

$$\sigma_{pq}(g) = I_{pq} g I_{pq}, \tag{2.3}$$

where $g \in U(p+q)$ and $I_{pq} = \left[\begin{smallmatrix} 1_p & \\ & -1_q \end{smallmatrix}\right]$. Accordingly the EF submanifold M_{pq} takes the following form:

$$M_{pq} = \{Q \in \mathrm{U}(p+q) \,|\, I_{pq} Q I_{pq} Q = \mathbb{1}\}. \tag{2.4}$$

In order to analyze the structure of M_{pq}, we parametrize G_{pq} in terms of projections. A point of $G_{pq}(\mathbb{C})$ is, by definition, a q-dimensional subspace of $\mathbb{C}^{p+q}$, and the latter is uniquely characterized by a Hermitian projection P of rank q, i.e., $\mathrm{Tr}\, P = q$. Our first observation is trivial.

Lemma 2.1: The relations

$$Q = I_{pq}(\mathbb{1} - 2P), \tag{2.5a}$$

$$P = \tfrac{1}{2}(\mathbb{1} - I_{pq}Q), \tag{2.5b}$$

define a bijection between the points $Q \in M_{pq}$ and projections P on *arbitrary* subspaces of $\mathbb{C}^{p+q}$. □

The proof is a straightforward verification of the equivalence between the relations $Q^\dagger Q = \mathbb{1}$, $I_{pq} Q I_{pq} Q = \mathbb{1}$ on one hand, and the relations $P^2 = P^\dagger = P$, on the other. Notice that $P = 0$ corresponds to $Q = I_{pq}$, $P = \mathbb{1}$ to $Q = -I_{pq}$.

It follows from Lemma 2.1 that the manifold $i_{pq}(G_{pq})$, identified with the set of all projections of rank q (the immersion i_{pq} is one-to-one in this case), cannot be identical to M_{pq}, which corresponds to projections of *all* ranks q'. But there is more.

Lemma 2.2: Let $p' + q' = p + q$ and define the map Φ: $U(p+q) \to U(p+q)$ by

$$\Phi(Q) = I_{p'q'} I_{pq} Q. \tag{2.6}$$

Then Φ is a diffeomorphism from M_{pq} onto $M_{p'q'}$ and the points $Q \in M_{pq}$, $\Phi(Q) \in M_{p'q'}$ correspond by (2.5) to the same projection P. □

This lemma, also straightforward, shows that M_{pq} cannot always be isomorphic to G_{pq}, since $\dim G_{pq} = 2pq \neq 2p'q' = \dim G_{p'q'}$ (unless $\{p',q'\} = \{p,q\}$). Let us come back to the discussion of Sec. I, setting $\mathfrak{g} = \mathfrak{u}(p+q)$, $\mathfrak{h} = \mathfrak{u}(p) \oplus \mathfrak{u}(q)$. Equation (1.9) for Ker Ψ_* becomes

$$X + (I_{pq}Q) X (I_{pq}Q)^{-1} = 0. \tag{2.7}$$

From this follows, as announced, that rank $\Psi|_{\mathbb{1}} = \dim \mathfrak{h}$ and rank $\Psi|I_{pq} = \dim \mathfrak{g}$. More generally, we can compute the rank of Ψ for each $Q = \pm I_{p'q'}$ $(p' + q' = p + q)$, and the result *does* vary with p', q'. Notice that in all cases $I_{p'q'} \in M_{pq}$: the rank of Ψ is not constant on M_σ. So the question remains: what is the structure of M_{pq}, and what are the properties of the Cartan immersion i_{pq}?

First we characterize the connected components $M_{pq}^{(k)}$ of M_{pq}. As in Ref. 3, we start from a point $Q \in M_{pq}^{(k)}$ and take a geodesic starting at Q, namely:

$$Q_t = Q e^{tX}, \quad t \in [0,1], \quad X \in Q\mathfrak{k}_Q. \tag{2.8}$$

This geodesic is entirely contained in $M_{pq}^{(k)}$, since the latter is totally geodesic. Correspondingly we get a one-parameter family of Hermitian projections

$$P_t = \tfrac{1}{2}(\mathbb{1} - I_{pq} Q e^{tX}), \quad t \in [0,1]. \tag{2.9}$$

Since $\mathrm{Tr}\, P_t$ is an integer and depends continuously on t, we get

$$\mathrm{Tr}\, P_t = \mathrm{Tr}\, P, \quad \forall\, t \in [0,1], \tag{2.10}$$

i.e., all projections corresponding to the points of the geodesic have the same rank. On the other hand, two projections of the same rank $P, \widetilde{P}$ are conjugated under $\mathrm{U}(p+q)$:

$$\widetilde{P} = U^\dagger P U, \quad U \in \mathrm{U}(p+q). \tag{2.11}$$

Since $\mathrm{U}(p+q)$ is connected and acts transitively on the subspaces of $\mathbb{C}^{p+q}$ of fixed dimensions, any two such projections are linked by a continuous family of projections, all of the same rank, and thus the corresponding points $\widetilde{Q}, Q$ of M_{pq} are connected by a continuous path in M_{pq} (broken geodesic) and in fact within the same connected component $M_{pq}^{(k)}$.

Thus each connected component $M_{pq}^{(k)}$ corresponds by (2.5) to *all* projections of a given rank q' $(0 \leqslant q' \leqslant p+q)$, i.e., it is the image of some Grassmann manifold $G_{p'q'}$ under the *corresponding* Cartan immersion $i_{p'q'}$. Combining this result with those of the general discussion above, we get the following theorem.

Theorem 2.3: We consider the complex Grassmann manifold $G_{pq}(\mathbb{C}) \cong \mathrm{U}(p+q)/\mathrm{U}(p) \times \mathrm{U}(q)$. The corresponding EF submanifold M_{pq} consists of the two isolated points $\{\pm I_{pq}\}$ and $(p+q-1)$ connected components $M_{pq}^{(k)}$, $k = 1, 2, \ldots, p+q-1$, such that the following conditions hold.

(i) For each k, $M_{pq}^{(k)}$ is a totally geodesic submanifold of $\mathrm{U}(p+q)$; it consists of all elements of the form $Q = I_{pq} g I_{p+q-k,k} g^{-1}$, $g \in \mathrm{U}(p+q)$, which correspond by the relations (2.5) to all projections of rank k in $\mathbb{C}^{p+q}$.

(ii) The Cartan immersion i_{pq} is a diffeomorphism of G_{pq} onto the identity component $(M_{pq})_0 = M_{pq}^{(q)}$.

(iii) More generally, each component $M_{pq}^{(k)}$ is the image of the Grassmann manifold $G_{p+q-k,k}$ under the diffeomorphism $\Phi^{(k)} \circ i_{p+q-k,k}$, where the map $\Phi^{(k)}$: $\mathrm{U}(p+q) \to \mathrm{U}(p+q)$, defined by $\Phi^{(k)}(Q) = I_{pq} I_{p+q-k,k} Q$, is a diffeomorphism of $M_{p+q-k,k}$ onto M_{pq}.

(iv) Finally, the components $M_{pq}^{(k)}$ and $M_{pq}^{(p+q-k)}$ are diffeomorphic to each other, under the map $Q \leftrightarrow -Q$, equivalent to $P \leftrightarrow \mathbb{1} - P$ under (2.5). □

To give an example, take $G_{31} = \mathrm{U}(4)/\mathrm{U}(3) \times \mathrm{U}(1)$ and $G_{22} = \mathrm{U}(4)/\mathrm{U}(2) \times \mathrm{U}(2)$. The corresponding Cartan immersions i_{31} and i_{22} map both manifolds into $\mathrm{U}(4)$, and one has $i_{31}(G_{31}) = (M_{31})_0 = M_{31}^{(1)}$, $i_{22}(G_{22}) = (M_{22})_0 = M_{22}^{(2)}$. The respective EF submanifolds M_{31}, M_{22} have the following structure, where the connected components correspond to each other via the diffeomorphism Φ: $\mathrm{U}(4) \to \mathrm{U}(4)$ given by $\Phi(Q) = I_{22} I_{31} Q$, as indicated in Lemma 2.2,

$$\begin{array}{lccccc}
M_{31} = & \{I_{31}\} \cup & [(M_{31})_0 = M_{31}^{(1)}] \cup & M_{31}^{(2)} & \cup\, M_{31}^{(3)} \cup & \{-I_{31}\} \\
& \downarrow \Phi & \downarrow \Phi & \downarrow \Phi & \downarrow \Phi & \downarrow \Phi \\
M_{22} = & \{I_{22}\} \cup & M_{22}^{(1)} & \cup\, [(M_{22})_0 = M_{22}^{(2)}] & \cup\, M_{22}^{(3)} \cup & \{-I_{22}\}.
\end{array} \tag{2.12}$$

For $k = 1, 2, 3$, the components $M_{31}^{(k)}$ and $M_{22}^{(k)}$ are diffeomorphic to the set $P^{(k)}$ of k planes in $\mathbb{C}^4$. For instance, the unit matrix $\mathbb{1}_4 \in (M_{31})_0$ and its image $\Phi(\mathbb{1}) = I_{22} I_{31} \in M_{22}^{(1)}$ correspond to the rank 1 projection $[{}^{0_3}{}_{1}]$, whereas the same unit matrix $\mathbb{1}_4$ taken in $(M_{22})_0$ and its image $\Phi^{-1}(\mathbb{1})$, correspond to the rank 2 projection $[{}^{0_2}{}_{1_2}]$, each time according to Eq. (2.5b).

The conclusion of the analysis is that the constraints

$Q^\dagger Q = \mathbb{1}$ (subgroup) and $I_{pq} Q I_{pq} Q = \mathbb{1}$ (quotient) are not sufficient to guarantee that the point $Q \in U(p+q)$ belongs to the image of the Cartan immersion $i_{pq}: G_{pq} \to U(p+q)$. One needs the additional constraint rank $P = q$, where P is the projection corresponding to Q, i.e., $P = \frac{1}{2}(\mathbb{1} - I_{pq} Q)$. Obviously this fact has important consequences for the analysis of nonprincipal σ models, which is mostly based on the result of Eichenherr and Forger[3] described in Sec. I. We will analyze in Sec. VI the implications of the additional constraint for the HSS method.[11,12]

Before proceeding to the noncompact case, let us describe briefly the alternative parametrization (2.2): $G_{pq}(\mathbb{C}) = SU(p+q)/S(U(p)\times U(q))$. The involution (2.3) remains the same, but we have a new Cartan immersion

$$\tilde{i}_{pq}: G_{pq}(\mathbb{C}) \to SU(p+q) \tag{2.13}$$

and a new EF submanifold

$$\widetilde{M}_{pq} = M_{pq} \cap SU(p+q) = \{Q \in M_{pq} \mid \det Q = 1\}. \tag{2.14}$$

According to (2.5a), we have

$$\det Q = \det I_{pq} \det(\mathbb{1} - 2P) = (-1)^{q + \operatorname{rank} P}, \tag{2.15}$$

since $\det(\mathbb{1} - 2P) = \det(\mathbb{1} - 2P_{\text{diag}})$, where

$$P_{\text{diag}} = U^\dagger P U = \begin{bmatrix} \mathbb{0} & \\ & \mathbb{1}_{\operatorname{rank} P} \end{bmatrix}. \tag{2.16}$$

Then the whole analysis may be repeated, with the following modifications: (i) the map Φ of Lemma 2.2 is a diffeomorphism from M_{pq} onto $M_{p'q'}$ only if $p' + q' = p + q$ and $|q - q'|$ is even; (ii) $I_{pq} \in SU(p+q)$ iff q is even and $-I_{p+q} \in SU(p+q)$ iff p is even; and (iii) the set $\widetilde{M}_{pq}$ contains only those components $M_{pq}^{(k)}$ for which $k \equiv q \pmod 2$, corresponding to projections of rank k. Combining these remarks with Theorem 2.3, we get the following structure for $\widetilde{M}_{pq}$:

$$\widetilde{M}_{pq} = \bigcup_k \widetilde{M}_{pq}^{(k)} \cup (\widetilde{M}_{pq})_{\text{isol}}, \tag{2.17}$$

where

(1) *for q even:* $\widetilde{M}_{pq}^{(k)} \simeq G_{p+k-2k,\,2k}$,

p *even:* $k = 1,2,\ldots,\frac{1}{2}(p+q) - 1$,

$(\widetilde{M}_{pq})_{\text{isol}} = \{I_{pq}, -I_{pq}\}$,

p *odd:* $k = 1,2,\ldots,\frac{1}{2}(p+q-1)$, $(\widetilde{M}_{pq})_{\text{isol}} = \{I_{pq}\}$;

(2) *for q odd:* $\widetilde{M}_{pq}^{(k)} \simeq G_{p+k-2k+1,\,2k-1}$,

p *even:* $k = 1,2,\ldots,\frac{1}{2}(p+q-1)$,

$(\widetilde{M}_{pq})_{\text{isol}} = \{-I_{pq}\}$,

p *odd:* $k = 1,2,\ldots,\frac{1}{2}(p+q)$, $(\widetilde{M}_{pq})_{\text{isol}} = \varnothing$,

Taking again the example of G_{31} as above, we get

$$\begin{array}{lccc} M_{31} = & [(M_{31})_0 = i_{31}(G_{31})] \cup & & M_{31}^{(3)}, \\ & \downarrow \Phi & & \downarrow \Phi \\ M_{13} = & M_{13}^{(1)} & \cup & [(M_{13})_0 = i_{13}(G_{13})], \end{array}$$

$$M_{22} = \{I_{22}\} \cup [(M_{22})_0 = i_{22}(G_{22})] \cup \{-I_{22}\}, \tag{2.18}$$

so that indeed M_{31} is diffeomorphic to M_{13}, but not to M_{22}.

III. GEOMETRY OF THE EF SUBMANIFOLD: NONCOMPACT GRASSMANNIANS

We turn now to the noncompact case, following closely the previous analysis. Again the relevant Grassmann manifold may be realized in two different ways,

$$G_{pi;qj}(\mathbb{C}) = U(p,q)/U(p-i,q-j)\times U(i,j) \tag{3.1}$$

$$= SU(p,q)/S(U(p-i,q-j)\times U(i,j)). \tag{3.2}$$

In the sequel we will use mostly the first realization.

The noncompact unitary group $U(p,q)$ consists of complex matrices g of order $(p+q)$ such that

$$g^\dagger t g = t, \tag{3.3}$$

where

$$t \equiv t_{pi;qj} = \begin{bmatrix} \mathbb{1}_{p-i} & & & \\ & -\mathbb{1}_{q-j} & & \\ & & \mathbb{1}_i & \\ & & & -\mathbb{1}_j \end{bmatrix}. \tag{3.4}$$

Correspondingly the involution defining $G_{pi;qj}(\mathbb{C})$, reads

$$\sigma(g) = IgI, \tag{3.5}$$

where

$$I \equiv I_{pi;qj} = \begin{bmatrix} \mathbb{1}_{p+q-i-j} & \\ & -\mathbb{1}_{i+j} \end{bmatrix} \equiv I_{p+q-i-j,\,i+j}. \tag{3.6}$$

This choice of t and I differs from the familiar one[7] used by HSS (by appropriate permutations of rows and columns) but is more convenient for our purposes in that it yields the *same* involution σ for all Grassmann manifolds (see Sec. V).

The manifold $G_{pi;qj}(\mathbb{C})$ consists of all vector subspaces of $\mathbb{C}^{p+q}$ of dimension $i+j$ and signature (i,j). The last condition means that only *nondegenerate* subspaces occur. As it is well known,[21] these subspaces are precisely those that correspond to orthogonal projections (with respect to the indefinite metric t of $\mathbb{C}^{p+q}$). Thus here too the natural parametrization of $G_{pi;qj}$ is in terms of projections (see Sec. V below).

There are two types of noncompact Grassmann manifolds: (i) for $i = 0, j = q$, $G_{po;qj}(\mathbb{C}) = U(p,q)/U(p)\times U(q)$ is a *Riemannian* symmetric space of noncompact type, dual to the compact RSS $G_{pq}(\mathbb{C})$ (it is in fact Hermitian and Kählerian), and (ii) for all other values of i, j, $G_{pi;qj}$ is a *pseudo-Riemannian* symmetric space (also pseudo-Hermitian, pseudo-Kählerian). The symmetric spaces of the second class have been classified by Berger[22] at the local level and by Shapiro[23] at the global level (the two are equivalent, for they are all simply connected). The complete list may be found in the monograph of Gilmore.[24] Nonlinear σ models on such pseudo-Riemannian symmetric spaces have been considered recently in the context of nonlinear dynamical systems[25] and in the reduction of Kaluza–Klein or supergravity theories.[26]

We proceed exactly as in the compact case. With the involution σ given in (3.5), the EF submanifold is

$$M_{pi;qj} = \{Q \in U(p,q) \mid IgIg = \mathbb{1}\}. \tag{3.7}$$

Lemma 2.1 is now replaced by the following lemma.

Lemma 3.1: The two relations

$$Q = I(\mathbb{1} - 2P), \tag{3.8a}$$

$$P = \tfrac{1}{2}(\mathbb{1} - IQ), \tag{3.8b}$$

where $I \equiv I_{pi;qj}$, define a bijection between the points $Q \in M_{pi;qj}$ and the projections P on nondegenerate subspaces of $\mathbb{C}^{p+q}$ of arbitrary dimension and signature. □

Again the proof is an immediate verification: the relations $Q^\dagger t Q = t$, $IQIQ = \mathbb{1}$ are equivalent to $P^2 = P$, $P^\dagger = tPt$.

The result corresponding to Lemma 2.2 is slightly different, in that both p and q are fixed, only i, j may vary.

Lemma 3.2: For any $0 \leqslant i' \leqslant p$, $0 \leqslant j' \leqslant q$, the map Φ: $\mathrm{U}(p,q) \to \mathrm{U}(p,q)$ defined by

$$\Phi(Q) = I_{pi';qj'} I_{pi;qj} Q \tag{3.9}$$

is a diffeomorphism from $M_{pi;qj}$ onto $M_{pi';qj'}$. □

Again Lemmas 3.1 and 3.2 show that $M_{pi;qj}$ is disconnected (except for very low dimensions) and its connected components are closed submanifolds, of different dimensions in general (however we still call $M_{pi;qj}$ the EF submanifold, as before). To characterize the connected components of $M_{pi;qj}$, we proceed as in the compact case. Two projections $P, \widetilde{P}$ in $\mathbb{C}^{p+q}$ are conjugated under $\mathrm{U}(p,q)$,

$$\widetilde{P} = U^\dagger P U, \quad U \in \mathrm{U}(p,q), \tag{3.10}$$

iff they project on (nondegenerate) subspaces of the same dimension and the same signature, by Witt's theorem.[27] Since $\mathrm{U}(p,q)$ is connected and its action on $\mathbb{C}^{p+q}$ has precisely for orbits the manifold of all subspaces of fixed dimension $i + j$ and signature (i,j), two such projections $P, \widetilde{P}$ may be linked by a continuous string of projections of the same type, and the corresponding points $Q, \widetilde{Q}$ are linked by a continuous path within a fixed connected component of $M_{pi;qj}$. Therefore each connected component $M^{(i'j')}_{pi;qj}$ corresponds to all projections of fixed rank $i' + j'$ and signature (i', j'), i.e., it is the image of the Grassmann manifold $G_{pi';qj'}$ under the corresponding Cartan immersion $i_{pi';qj'}$ and the map Φ of Lemma 3.2. In summary we state the following theorem.

Theorem 3.3: We consider the noncompact Grassmann manifold $G_{pi;qj}(\mathbb{C}) \cong \mathrm{U}(p,q)/\mathrm{U}(p-i, q-j) \times \mathrm{U}(i,j)$, $0 \leqslant i \leqslant p$, $0 \leqslant j \leqslant q$. The corresponding EF submanifold $M_{pi;qj}$ consists of two isolated points $\{\pm I_{pi;qj}\}$ and $(p+1) \times (q+1) - 2$ connected components $M^{(i',j')}_{pi;qj}$ $[0 \leqslant i' \leqslant p$, $0 \leqslant j' \leqslant q$, $(i',j') \neq (0,0),(p,q)]$. Each component $M^{(i',j')}_{pi;qj}$ is a totally geodesic submanifold of $\mathrm{U}(p,q)$, diffeomorphic to the Grassmann manifold $G_{pi',qj'}$ via the corresponding Cartan immersion $i_{pi';qj'}$, composed with the map Φ of Lemma 3.2. In particular the identity component $(M_{pi;qj})_0 = M^{(i,j)}_{pi;qj}$ is diffeomorphic to $G_{pi;qj}$ itself. Furthermore, the components $M^{(i',j')}_{pi;qj}$ and $M^{(p-i',q-j')}_{pi;qj}$ are diffeomorphic to each other under the map $Q \leftrightarrow -Q$, equivalent to $P \leftrightarrow \mathbb{1} - P$ for the corresponding projections in $\mathbb{C}^{p+q}$. □

Obviously the conclusions of the analysis are the same as in the compact case. For ensuring that a point $Q \in \mathrm{U}(p,q)$ belongs to the image of the Cartan immersion $i_{pi;qj}$: $G_{pi;qj} \to \mathrm{U}(p,q)$, one needs the two conditions specifying the EF manifold, namely $Q^\dagger t Q = t$ (subgroup) and $I_{pi;qj} Q I_{pi;qj} Q = \mathbb{1}$ (quotient), and in addition the two geometric constraints selecting the identity component of the EF submanifold $M_{pi;qj}$, namely rank $P = i + j$, sgn $P = (i,j)$, where $P = \tfrac{1}{2}(\mathbb{1} - IQ)$ is the projection in $\mathbb{C}^{p+q}$ corresponding to Q. As we will see in Sec. VI below this has dramatic consequences for σ models.

We conclude this section by examining briefly the alternative realization (3.2) of $G_{pi;qj}$. With an analysis entirely parallel to the one made in Sec. II, one gets the following modifications: (i) in Lemma 3.1, the condition $\det Q = 1$ is equivalent to rank $P \equiv (i+j) \bmod 2$; (ii) the map Φ of Lemma 3.2 is a diffeomorphism iff $i' + j' \equiv (i+j) \bmod 2$; (iii) $I_{pi;qj} \in \mathrm{SU}(p,q)$ iff $i + j$ is even, $-I_{pi;qj} \in \mathrm{SU}(p,q)$ iff $p + q - i - j$ is even; and (iv) the connected components of $M_{pi;qj}$ in $\mathrm{SU}(p,q)$ are indexed by i', j' as in Theorem 3.3, with the additional restriction $i' + j' \equiv (i+j) \bmod 2$.

To give an example, take $G_{31;21} = \mathrm{SU}(3,2)/S(\mathrm{U}(2,1) \times \mathrm{U}(1,1))$. Its EF submanifold $M_{31;21} \subset \mathrm{SU}(3,2)$ consists of (i) the isolated point $Q = I_{31;21} = [{}^{1_3}{}_{-1_2}]$, corresponding to $P = 0$; (ii) three connected components with rank $P = 2$, and signature $(i',j') = (2,0)$, $(1,1)$, $(0,2)$, respectively (among these $M^{(1,1)}_{31;21}$ is the identity component, image of $G_{31;21}$); and (iii) two connected components with rank $P = 4$ and signature $(i',j') = (2,2)$ and $(3,1)$, respectively. The component $M^{(0,2)}_{31;21}$ is diffeomorphic to $G_{30;32} = \mathrm{SU}(3,2)/S(\mathrm{U}(3) \times \mathrm{U}(2))$ and it is the only RSS that occurs.

IV. OTHER SYMMETRIC SPACES

The geometrical analysis of Secs. II and III applies to a variety of other symmetric spaces, with similar results. The EF submanifold M_σ is in general not connected and, moreover, the map i_σ is not always injective. We describe briefly several cases, compact or not.

A. Real Grassmann manifolds

Restricting the analysis of Sec. II from $\mathbb{C}^n$ to $\mathbb{R}^n$, one gets two different symmetric spaces of compact type, corresponding to the involution σ_{pq} given in (2.3) (see Refs. 9 or 28 for more details):

the real Grassmann manifold,

$$G_{pq}(\mathbb{R}) = \mathrm{O}(p+q)/\mathrm{O}(p) \times \mathrm{O}(q) = \mathrm{SO}(p+q)/S(\mathrm{O}(p) \times \mathrm{O}(q)); \tag{4.1}$$

the oriented real Grassmann manifold,

$$\widetilde{G}_{pq}(\mathbb{R}) = \mathrm{SO}(p+q)/\mathrm{SO}(p) \times \mathrm{SO}(q). \tag{4.2}$$

As it is well known $\widetilde{G}_{pq}$ is a twofold covering of G_{pq}, and, as a consequence, the Cartan immersion $\tilde{i}_{pq}$: $\widetilde{G}_{pq} \to \mathrm{SO}(p+q)$ is two-to-one for $(p+q)$ even, whereas i_{pq}: $G_{pq} \to \mathrm{SO}(p+q)$ is injective (obviously $\tilde{i}_{pq}$ factorizes through G_{pq}). Again points of G_{pq} are represented by orthogonal projections of rank q in $\mathbb{R}^{p+q}$ and $\mathrm{SO}(p+q)$ acts transitively on G_{pq}. Exactly as in the complex case discussed in Sec. II, i_{pq} is a diffeomorphism from G_{pq} onto the identity component $(M_{pq})_0$, but $\tilde{i}_{pq}$ is only a *local* diffeomorphism onto $(M_{pq})_0$ (see Sec. I and the Erratum to Ref. 3). We omit the details.

In the noncompact case, one gets only one kind of symmetric space, namely,

$$G_{pi;qj}(\mathbb{R}) = \mathrm{SO}_0(p,q)/\mathrm{SO}_0(p-i, q-j) \times \mathrm{SO}_0(i,j), \tag{4.3}$$

and the analysis proceeds exactly as in Sec. III. Notice that none of those spaces is (semi)Kählerian, except for $p+q=3$ because of the isomorphisms

$$G_{21}(\mathbb{R}) \simeq G_{11}(\mathbb{C}), \quad G_{20;11}(\mathbb{R}) \simeq G_{10;11}(\mathbb{C}). \qquad (4.4)$$

B. Symplectic or quaternionic Grassmann manifolds

A third type of Grassmann manifold is obtained if one replace the unitary groups in Secs. II and III by their symplectic subgroup, namely, the following cases.

(i) In the *compact* case,

$$G_{pq}(\mathbb{H}) = \mathrm{Sp}(p+q)/\mathrm{Sp}(p)\times \mathrm{Sp}(q) \qquad (4.5)$$

with the involution $\sigma_{2p,2q}$ given by Eq. (2.3).

(ii) In the *noncompact* case,

$$G_{pi;qj}(\mathbb{H}) = \mathrm{Sp}(p,q)/\mathrm{Sp}(p-i,q-j)\times \mathrm{Sp}(i,j) \qquad (4.6)$$

with the involution $\sigma_{2p,2i;2q,2j}$ of Eq. (3.5). The important point is that $\mathrm{Sp}(p+q)$ [resp. $\mathrm{Sp}(p,q)$], is a subgroup of $\mathrm{SU}(2p+2q)$ [resp. $\mathrm{SU}(2p,2q)$], defined by an additional constraint. Indeed a matrix of order $(2p+2q)$ belongs to $\mathrm{Sp}(2p+2q)$ [resp. $\mathrm{Sp}(2p,2q)$], if it verifies the relations

$$g^\dagger t g = t, \qquad (4.7a)$$

$$g^T J g = J. \qquad (4.7b)$$

The metric t in (4.7a) equals $\mathbb{1}$ in the compact case and $t_{2p,2i;2q,2j}$ as given by Eq. (3.4), in the noncompact one. In both cases, J denotes the matrix

$$J \equiv J_{pi;qj} = \begin{bmatrix} J_{p-i} & & & \\ & J_{q-j} & & \\ & & J_i & \\ & & & J_j \end{bmatrix},$$

$$\text{with } J_n = \begin{bmatrix} & \mathbb{1}_n \\ -\mathbb{1}_n & \end{bmatrix}, \qquad (4.8)$$

where one takes $i=0$, $j=q$ for the compact case, i.e., $J_{pq} \equiv \left[\begin{smallmatrix} J_p & \\ & J_q \end{smallmatrix}\right]$.

The connections (2.7) and (3.9) between points of the Grassmannian and projections are still valid, because the additional constraint (4.7b), rewritten as $Q = -tJ\bar{Q}Jt$, is equivalent to the relation $P = -tJ\bar{P}Jt$. Then the whole analysis goes through, because the involutions in the symplectic case are exactly those of the complex case. Indeed, let $M_{2p;2q}$ denote the complex EF submanifold given in Eq. (2.4) and define the symplectic one as follows:

$$M^s_{pq} = \{Q \in \mathrm{Sp}(p+q) \,|\, I_{2p,2q} Q I_{2p,2q} Q = \mathbb{1}\}. \qquad (4.9)$$

Then one has

$$M^s_{pq} = M_{2p,2q} \cap \mathrm{Sp}(p+q). \qquad (4.10)$$

Since the action of $\mathrm{Sp}(p+q)$ on $G_{pq}(\mathbb{H})$ is transitive, the same relation holds between the connected components of the corresponding EF submanifolds

$$(M^s_{pq})^{(k)} = M^{(k)}_{2p,2q} \cap \mathrm{Sp}(p+q), \qquad (4.11)$$

and so the analysis of Sec. II may be repeated word for word. The same is true in the noncompact case. Here again, the Cartan immersion is a diffeomorphism between the symplectic Grassmann manifold and the identity component of the corresponding EF submanifold, and each connected component of the latter is diffeomorphic to some symplectic Grassmann manifold.

C. Other examples

Thus, in all cases, a (connected) Grassmann manifold is *not* diffeomorphic to the corresponding EF submanifold, but only to the identity component of the latter (except for trivial, low-dimensional examples). Does this property extend to arbitrary symmetric spaces? Going through the table,[11,24] no further instances are found of isomorphisms between the EF submanifolds of two different symmetric spaces with the same group of isometries, such as the maps Φ of Lemmas 2.2 and 3.2, so we have to proceed differently.

Comparing Eqs. (1.6) and (1.7), we see that the EF submanifold M_σ of the space G/H is connected if, for every $Q \in M_\sigma$, there exists $g \in G$ such that $Q = \sigma(g)g^{-1}$.

We take first the AI space of *noncompact* type $\mathrm{SL}(n,\mathbb{R})/\mathrm{SO}(n)$, for which $\sigma(g) = (g^T)^{-1}$. Thus $Q \in M_\sigma$ is a real symmetric matrix of determinant 1, and $Q \in (M_\sigma)_0$ if $Q = gg^T$ for some $g \in \mathrm{SL}(n,\mathbb{R})$, i.e., if it is strictly positive definite. However, if $Q \in M_\sigma$, $-Q \in M_\sigma$ also (for n even), but $-Q$ is not positive definite; hence for n even, M_σ contains at least two disjoint components (diffeomorphic to each other). Exactly the same argument applies to all remaining spaces of type III, namely $\mathrm{SU}^*(2n)/\mathrm{Sp}(n)$ (AII), $\mathrm{SO}^*(2n)/\mathrm{U}(n)$ (DIII), and $\mathrm{Sp}(n,\mathbb{R})/\mathrm{U}(n)$ (CI), for all n, and also to the spaces of type IV, for n even. In each case, the EF submanifold has at least two disjoint components, diffeomorphic to each other by the relation $Q \leftrightarrow -Q$, whereas $(M_\sigma)_0$ consists of strictly positive definite matrices.

For the *compact* spaces of type I, however, the situation is different. First of all, the argument above fails. Take for instance the (AI) space $\mathrm{SU}(n)/\mathrm{SO}(n)$ for n even. Then $Q \in (M_\sigma)_0$ means $Q = gg^T$, with $Q,g \in \mathrm{SU}(n)$; such a matrix Q is not positive definite and indeed $-Q$ is of the same type, since $-Q = g'g'^T$, with $g' = ig\left[\begin{smallmatrix} & 1 \\ 1 & \end{smallmatrix}\right] \in \mathrm{SU}(n)$. Similarly for the remaining spaces of type I, namely the series AII, DIII, and CI. But there is more. For the space $\mathrm{SU}(n)/\mathrm{SO}(n)$, it can be shown explicitly[29] that a matrix $Q \in \mathrm{SU}(n)$ is symmetric iff it is of the form $Q = gg^T$, with $g \in \mathrm{SU}(n)$. So this symmetric space *is* diffeomorphic to its EF submanifold [although rank Ψ is not constant over $\mathrm{SU}(n)$, it is constant over the submanifold $M_\sigma = \Psi^{-1}(\mathbb{1})$]. The same is true[29] for the compact AII space $\mathrm{SU}(2n)/\mathrm{Sp}(n)$. For the remaining spaces, DIII and CI, the question is open, but we conjecture that these too are diffeomorphic to their EF submanifold.

In conclusion there remains a number of symmetric spaces for which the complete structure of the EF submanifold is unknown, the compact spaces DIII and CI, and the noncompact, non-Grassmannian spaces (type III and IV). This problem seems difficult; it is easy to characterize M_σ by the quotient constraints listed in Ref. 11, but usually not $i_\sigma(M)$.

V. PARAMETRIZATION OF GRASSMANNIAN σ MODELS IN TERMS OF PROJECTIONS

In view of the geometrical interpretation of Grassmann manifolds, it is natural to parametrize the corresponding σ

models in terms of projections, with the appropriate rank and signature. Indeed this is by now the standard procedure.[2–4,13–16,19] We will give in this section a unified presentation of this formulation, valid for all models, compact or not, i.e., leading to the same equations of motion. As recalled in the Introduction, the action (1.4) for the σ model valued in $M = G/H$ and the equation of motion may be reformulated in terms of a covariant derivative $D^{\mu}g$. For each Grassmannian model, $D^{\mu}g$ in turn may be expressed in terms of projections, as we will now see.

A. Complex Grassmannian models

We take first the compact case $G_{pq}(\mathbb{C})$. Writing the matrix $g \in SU(p+q)$ as $g = [Y;Z]$, where Y, Z have p and q columns, respectively, we see from the constraint $g^{\dagger}g = 1$ that the q columns of Z constitute an orthogonal basis of a q-dimensional subspace of $\mathbb{C}^{p+q}$, with associated projection

$$P = ZZ^{\dagger}. \tag{5.1}$$

Thus the projection P corresponds to a unique point of $G_{pq}(\mathbb{C})$, and will be taken as basic field.

In the noncompact case $G_{pi;qj}(\mathbb{C})$, we write $g \in SU(p,q)$ as $g = [Y;Z]$ where Z contains the last $(i+j)$ columns of g. The constraint $g^{\dagger}tg = t$, where $t = t_{pi;qj}$, Eq. (3.4), then shows that the columns of Z are an orthogonal (with respect to the metric t) basis of a $(i+j)$-dimensional subspace of $\mathbb{C}^{p+q}$, with signature (i,j). The corresponding projection is

$$P = Zt_Z Z^{\dagger}t, \tag{5.2}$$

where t_Z is the metric in the $(i+j)$-dimensional subspace

$$t_Z = Z^{\dagger}tZ = I_{ij}. \tag{5.3}$$

Indeed, it is straightforward to verify the relations $P^2 = P = tP^{\dagger}t$. The projection P thus corresponds uniquely to a point of $G_{pi;qj}$, and will henceforth be taken as the basic field of the σ model. Finally notice that replacing t by $\mathbb{1}$ in (5.2) and (5.3) yields $t_Z = \mathbb{1}$ and $P = ZZ^{\dagger}$, i.e., the formulas for the compact case. So it is enough to write the noncompact case. In terms of P, the covariant derivative reads[18,19]

$$D_{\mu}Z = (\mathbb{1} - Zt_Z Z^{\dagger}t)\partial_{\mu}Z = (\mathbb{1} - P)\partial_{\mu}Z, \tag{5.4}$$

the Lagrangian becomes

$$\mathscr{L} = \tfrac{1}{2}\operatorname{Tr}\{(D_{\mu}Z)^{\dagger}t(D^{\mu}Z)t_Z\} = \operatorname{Tr}(\partial_{\mu}P)(\partial^{\mu}P) \tag{5.5}$$

and the equation of motion

$$D_{\mu}D^{\mu}Z - Zt_Z(D_{\mu}Z)^{\dagger}t(D^{\mu}Z) = 0$$

is equivalent to

$$[\Box P, P] = 0, \quad \Box = \partial_{\mu}\partial^{\mu} = \partial_0^2 \pm \partial_1^2. \tag{5.6}$$

B. Real Grassmannian models

All the previous formulas remain valid in the real case, just by taking restrictions from $\mathbb{C}^{p+q}$ to $\mathbb{R}^{p+q}$ everywhere.

C. Symplectic Grassmannian models

In the symplectic case, the projections must verify the additional constraint

$$P = -tJ\bar{P}Jt \tag{5.7}$$

as shown in Sec. IV B. To construct such a projection, we proceed exactly as in Sec. V A above. Given $g \in SU(2p,2q)$, we write it again as $g = [Y,Z]$, where Z consists of the last $2(i+j)$ columns of g. We further decompose it as $Z = [Z_{++};Z_{+-};Z_{-+};Z_{--}]$, where the blocks $Z_{+\pm}$ and $Z_{-\pm}$ have, respectively, i and j columns, according to the block decomposition of t and J. The constraints (4.7b) [resp. (4.7a)] imply the following relations on the matrices $Z_{\alpha\beta}(\alpha,\beta = \pm)$:

$$Z_{\alpha -} = -Jt\,\overline{Z_{\alpha +}}, \tag{5.8a}$$

$$Z^{\dagger}_{\alpha\beta}tZ_{\alpha\beta} = t_{\alpha}, \quad \text{where } t_+ = \mathbb{1}_i, \quad t_- = -\mathbb{1}_j. \tag{5.8b}$$

Then the matrices

$$P_{\alpha\beta} = Z_{\alpha\beta}t_{\alpha}Z^{\dagger}_{\alpha\beta}t \quad (\alpha,\beta = \pm) \tag{5.9}$$

are four, mutually orthogonal, Hermitian projections,

$$P_{\alpha\beta}P_{\alpha'\beta'} = \delta_{\alpha\alpha'}\delta_{\beta\beta'}P_{\alpha\beta} = tP^{\dagger}_{\alpha\beta}t. \tag{5.10}$$

Defining further $P_{\beta} = P_{+\beta} + P_{-\beta}$, we get two projections of rank $(i+j)$ and signature (i,j), whose images are mutually conjugate subspaces. They also verify the following relation:

$$P_{\pm} = -tJ\,\overline{P_{\mp}}\,Jt. \tag{5.11}$$

Consider now the projection

$$P = P_+ + P_- = \sum_{\alpha,\beta = \pm} P_{\alpha\beta}. \tag{5.12}$$

This is the field of the noncompact symplectic Grassmannian model. It projects on the appropriate subspace and verifies all the constraints, in particular (5.7), which follows trivially from (5.11). To get the corresponding result for the compact case, it suffices again to replace everywhere t by $\mathbb{1}$, which yields $t_{\pm} = \mathbb{1}_q$ as before. Finally, as in the previous cases, the equation of motion reduces to Eq. (5.6).

VI. THE BÄCKLUND TRANSFORMATION METHOD REVISITED

The Bäcklund transformation method introduced by Harnad, Saint-Aubin, and Shnider[11,12] is an elegant and powerful technique for constructing multisoliton solutions of $(1+1)$-dimensional integrable systems, in particular Minkowskian RSS-valued σ models.

In this section, we will show that the HSS method applies to Euclidean models as well, with minor modifications. On the other hand, we shall also reexamine its validity in the light of the previous discussion.

The systems considered are of the Zakharov–Mikhailov–Shabat[13,30,31] (ZMS) type, namely

$$\psi_{\xi} = U\psi, \quad \psi_{\eta} = V\psi, \tag{6.1}$$

where $U(\lambda,\xi,\eta)$, $V(\lambda,\xi,\eta)$, and $\psi(\lambda,\xi,\eta)$ are $n \times n$ matrix functions of ξ, $\eta \in \mathbb{C}$, depending on a complex parameter λ, with ψ invertible and U, V meromorphic. The ZMS "dressing method" consists in obtaining a new solution $(\tilde{U},\tilde{V},\tilde{\psi})$ of (6.1) from a given one (U,V,ψ) with help of a "dressing matrix" $\chi(\lambda,\xi,\eta) \in SL(n,\mathbb{C})$; in particular,

$$\tilde{\psi} = \chi\psi. \tag{6.2}$$

The matrix χ and its inverse χ^{-1} are assumed to be meromorphic in λ, with simple poles $\{\lambda_i\}_{i=1,\dots,K}$, $\{\mu_j\}_{j=1,\dots,K}$, respectively, and normalized to $\mathbb{1}$ at $\lambda=\infty$, so that

$$\chi(\lambda)=\mathbb{1}+\sum_{i=1}^{K}\frac{Q_i}{\lambda-\lambda_i},\qquad \chi^{-1}(\lambda)=\mathbb{1}+\sum_{j=1}^{K}\frac{R_j}{\lambda-\mu_j}. \tag{6.3}$$

The idea of HSS is to determine the dressing matrix χ in terms of a $(nK\times nK)$ matrix $M(\xi,\eta)$, called the soliton correlation matrix, which satisfies a system of Riccati equations [Ref. 12, Eq. (2.7)]. The integrability condition of that system is that the following $\mathfrak{gl}(2nK,\mathbb{C})$-valued one-form have zero curvature:

$$\omega=-\begin{pmatrix}p_+ & 0\\ r_+ & s_+\end{pmatrix}d\xi-\begin{pmatrix}p_- & 0\\ r_- & s_-\end{pmatrix}d\eta, \tag{6.4}$$

where $p_\pm, r_\pm, s_\pm$, the coefficients in the Riccati system, are given in terms of the values of U and V at the poles $\{\lambda_i,\mu_j\}$. For the particular case of the principal σ model on $\mathrm{SL}(n,\mathbb{C})$, one has

$$U=g_\xi g^{-1}/(1+\lambda),\quad V=g_\eta g^{-1}/(1-\lambda), \tag{6.5}$$

where

$$g(\xi,\eta)=\psi(0,\xi,\eta). \tag{6.6}$$

Thus the BT transformation (6.2) reads in this case

$$\tilde g=\left(\mathbb{1}-\sum_{i=1}^{K}\frac{Q_i}{\lambda_i}\right)g. \tag{6.7}$$

Up to this point, the treatment is valid for arbitrary complex $\xi,\eta\in\mathbb{C}$, in particular for Minkowskian models (with light cone coordinates $\xi=x+t$, $\eta=x-t$) and Euclidean models (with complex coordinates $\xi=x+it$, $\eta=\bar\xi=x-it$). But a distinction enters when one performs the reduction from $\mathrm{SL}(n,\mathbb{C})$ to various subgroups and quotients. Following HSS, such a reduction is defined by a linear fractional transformation

$$s:\ \lambda\mapsto\frac{a\lambda+b}{c\lambda+d},\quad \begin{pmatrix}a & b\\ c & d\end{pmatrix}\in\mathrm{SL}(2,\mathbb{C}), \tag{6.8}$$

and an automorphism σ of $\mathrm{GL}(n,\mathbb{C})$. The particular reduction chosen is expressed by a set of invariance conditions on ψ, U, V [Ref. 12, Eq. (3.9)]. When the map σ_* is linear ($\sigma=\sigma_1$ or σ_3 in the notations of HSS), those conditions remain unchanged for a Euclidean model. But since ξ and $\eta=\bar\xi$ are interchanged under complex conjugation, the matrices U and V must be interchanged too when σ_* is antilinear ($\sigma=\sigma_2$ or σ_4), i.e., the invariance conditions become

$$\psi(s(\bar\lambda))=f\sigma\psi(\lambda), \tag{6.9a}$$

$$U(s(\bar\lambda))=f\sigma_*[V(\lambda)]f^{-1}+f_\xi f^{-1}, \tag{6.9b}$$

$$V(s(\bar\lambda))=f\sigma_*[U(\lambda)]f^{-1}+f_\eta f^{-1}, \tag{6.9c}$$

where the gauge function f is fixed by evaluating (6.9a) at $\lambda=\infty$ as in HSS.

However the constraint on the dressing matrix remains the same as in the Minkowskian case, namely Eq. (3.13) of Ref. 12,

$$\sigma(\chi(\lambda))=\bar f^{-1}\chi(s(\bar\lambda))f, \tag{6.10}$$

and therefore the constraints on the correlation matrix $M(\xi,\eta)$ [Eq. (3.14) of Ref. 12], which are equivalent to (6.10), remain the same too. Thus the only point to check is whether those constraints on $M(\xi,\eta)$ are compatible with the Riccati system, i.e., whether a solution $M(\xi,\eta)$ satisfies the constraints at all points (ξ,η) if the initial condition $M(\xi_0,\eta_0)$ does. The compatibility is expressed in terms of the one-form ω defined in (6.4), namely,

$$\text{for } \sigma=\sigma_2,\ \ d\tilde L+\omega\tilde L-\tilde L\bar\omega=0,\qquad \text{for } \sigma=\sigma_4,\ \ dS-\omega^\dagger S-S\omega=0, \tag{6.11}$$

where $\tilde L$, S are matrices depending on s and the positions of the poles $\{\lambda_i,\mu_j\}$ of $\chi^{\pm1}(\lambda)$, given in Eqs. (3.16) and (3.18) of Ref. 12. Using the fact that complex conjugation exchanges ξ and η in ω, it is straightforward to show, as in Ref. 12, Theorem 3.2, that condition (6.11) indeed holds. Thus the analysis of HSS goes through completely for Euclidean models too.

An important consequence of the interchange between U and V in the constraints (6.9) is that the map s must interchange the poles of U and V. For instance, in the σ models, where U and V are given in (6.7), one must have $s(1)=-1$, $s(-1)=1$. This fact entails a modification of the minimal set of poles that must be chosen for χ in the reduction process. For the Euclidean complex Grassmannian σ model for example, the minimal set of poles is $(\lambda, 1/\lambda)$ for χ and $(-\bar\lambda,-1/\bar\lambda)$ for χ^{-1} (this fact was already observed by Sasaki[16]).

We may also notice that the HSS analysis is not limited to the case of *Riemannian* symmetric spaces G/H, it is applicable to pseudo-Riemannian ones as well. Indeed the metric on G/H is never used, only the constraints, and these we have discussed in the general case.

After these considerations, however, there remains an important point. Namely, the HSS analysis, as widely applicable as it is, is not sufficient *per se* for the σ model on an arbitrary symmetric space $M=G/H$. Indeed, following Ref. 3, the HSS reduction technique for obtaining solutions of the G/H model consists in taking solutions of the principal $\mathrm{SL}(n,\mathbb{C})$ model and subjecting them to appropriate constraints of subgroup and quotient type, where the latter guarantees that the solution belongs to the EF submanifold M_σ. But, as we know, this is not enough, an additional constraint is needed, forcing the solution to live in the identity component $(M_\sigma)_0$, diffeomorphic to G/H. Thus we have to check whether the HSS method verifies the topological constraint, i.e., whether the solution $\tilde g=\chi(0)g$ belongs to $(M_\sigma)_0$ whenever the initial data g does. In its full generality, this problem seems difficult and we have been unable to solve it. However, for Grassmannian models, we can give an answer. In terms of projections the BT solution reads, taking (6.7) into account (I is the matrix relevant for the model treated, see Secs. II–IV,

$$\tilde P=\tfrac12(\mathbb{1}-I\tilde g)=P+\frac12 I\sum_{i=1}^{K}\frac{Q_i g}{\lambda_i}. \tag{6.12}$$

First we check the rank,

$$\mathrm{Tr}\,\widetilde{P} = \mathrm{Tr}\,P + \frac{1}{2}\,\mathrm{Tr}\Big(\sum_i \frac{1}{\lambda_i} IQ_i g\Big)$$

and show that the second term indeed vanishes. For this we need some more notation from HSS. Let $q_i = \mathrm{rank}\,Q_i$, $r_i = \mathrm{rank}\,R_i$. Then the residue Q_i may be written as

$$Q_i = \Big(\sum_{j=i}^{K} H_j \gamma_{ji}\Big) F_i^\dagger, \tag{6.13}$$

where $H_i \in \mathbb{C}^{n\times r_i}$, $F_i \in \mathbb{C}^{n\times q_i}$ are solutions of a linear system. The matrix $\gamma = (\gamma_{ij})$, $i,j = 1,...,K$, is an invertible matrix, whose inverse $\gamma^{-1} = \Gamma = (\Gamma_{ij})$ consists of the following $n\times n$ blocks:

$$\Gamma_{ij} = F_i^\dagger H_j/(\lambda_i - \mu_j), \quad \text{if } \lambda_i \neq \mu_j \ \ (\text{thus } i\neq j). \tag{6.14}$$

For $\lambda_i = \mu_j$, the diagonal elements Γ_{ii} are defined appropriately (see Ref. 11, Theorem 4.2), and in that case, $F_i^\dagger H_i = 0$.

For all three cases of Grassmannian σ models, the poles of χ and χ^{-1} come in pairs $(\lambda_i, \lambda_{\hat{\imath}} \equiv 1/\lambda_i)$ [resp. $(\mu_i, \mu_{\hat{\imath}} \equiv 1/\mu_i)$] and the reduction is associated to the quotient automorphism $\sigma_-(g) = IgI$, as explained in Secs. II–IV. Notice that σ_- does not distinguish between Euclidean and Minkowskian kinematics, and so the argument is valid for both types of models (the distinction between the two is given by the relation between the poles of χ and those of χ^{-1}: $\mu_i = \bar{\lambda}_i$ in the Minkowskian case, $\mu_i = -\bar{\lambda}_i$ in the Euclidean one). Accordingly, the constraints imposed by the reduction on the residues Q_i, R_j read

$$(g^\dagger)^{-1} IF_i = F_{\hat{\imath}} \Lambda_i, \tag{6.15a}$$

$$gIH_i = H_{\hat{\imath}} \Xi_i, \tag{6.15b}$$

$$\lambda_i \mu_j \Gamma_{ij} = -\Lambda_i^\dagger \Gamma_{\hat{\imath}\hat{\jmath}} \Xi_j, \tag{6.15c}$$

where $\Lambda_i \in \mathrm{GL}(q_i,\mathbb{C})$, $\Xi_i \in \mathrm{GL}(r_i,\mathbb{C})$ are constant invertible matrices.

Using these notations, we prove our statement

$$\mathrm{Tr}\Big(\sum_i \frac{1}{\lambda_i} IQ_i g\Big) = \mathrm{Tr}\Big(\sum_{ij} \frac{1}{\lambda_i} gIH_j \gamma_{ji} F_i^\dagger\Big)$$

$$= \mathrm{Tr}\Big(\sum_{ij} \frac{1}{\lambda_i} H_{\hat{\jmath}} \Xi_j \gamma_{ji} F_i^\dagger\Big) \quad \text{by (6.15b)}$$

$$= \mathrm{Tr}\Big(\sum_{ij} \frac{1}{\lambda_i} (F_i^\dagger H_{\hat{\jmath}}) \Xi_j \gamma_{ji}\Big)$$

$$= \mathrm{Tr}\Bigg(\sum_{\substack{ij \\ \lambda_i \neq \mu_{\hat{\jmath}}}} \frac{\lambda_i - \mu_{\hat{\jmath}}}{\lambda_i} \gamma_{ji} \Gamma_{i\hat{\jmath}} \Xi_j\Bigg)$$

(the other terms vanish)

$$= \mathrm{Tr} \sum_{\substack{ij \\ \lambda_i \neq \mu_{\hat{\jmath}}}} \Big(\gamma_{ji} \Gamma_{i\hat{\jmath}} \Xi_j + \frac{1}{\lambda_i^2} \gamma_{ji} (\Lambda_i^\dagger)^{-1} \Gamma_{\hat{\imath}j}\Big).$$

Reinstating the missing terms $(i = \hat{\jmath},\ \lambda_i = \mu_{\hat{\jmath}})$, this yields

$$\cdots = \mathrm{Tr}\Big(\sum_{ij} \gamma_{ji} \Gamma_{i\hat{\jmath}} \Xi_j + \sum_{ij} \frac{1}{\lambda_i^2} \Gamma_{\hat{\imath}j} \gamma_{ji} (\Lambda_i^\dagger)^{-1}\Big)$$

$$- \mathrm{Tr} \sum_{\substack{i \\ \lambda_i = \mu_{\hat{\jmath}}}} \Big(\gamma_{\hat{\imath}i} \Gamma_{ii} \Xi_{\hat{\imath}} + \frac{1}{\lambda_i^2} \Gamma_{\hat{\imath}\hat{\imath}} \gamma_{\hat{\imath}i} (\Lambda_i^\dagger)^{-1}\Big)$$

and this is indeed 0; each term on the first line vanishes, since $\Sigma_i \gamma_{ji} \Gamma_{i\hat{\jmath}} = \delta_{j\hat{\jmath}} = 0$, and so does each bracket on the second line, by (6.15c), with $\hat{\jmath} = i$, $\mu_{\hat{\jmath}} = \lambda_i$.

Thus the BT (6.9) preserves the rank of the projection. This means that, for all σ models on Grassmann manifolds of *compact* type, $\tilde{g}$ lies in $(M_\sigma)_0$ if g does, in other words the HSS method is complete. However we cannot conclude in the noncompact case: the BT does indeed conserve the rank of the projection, but it must also conserve the signature, and this is open to question.

In conclusion the BT method of Harnad, Saint-Aubin, and Shnider[11,12] is fully justified for the σ models, either Minkowskian or Euclidean, with values in the following symmetric spaces: (i) those for which the EF submanifold M_σ is connected, equivalently for which the Cartan immersion maps G/H *onto* M_σ; for instance, the *compact* spaces of the series AI and AII, namely SU(n)/SO(n) and SU($2n$)/Sp(n); and (ii) all Grassmann manifolds of *compact* type (the AIII, BDI, and CII series, in their compact realization). For the other cases, the validity of the method remains in question. This applies, in particular, to the space SL(n,ℝ)/SO(n), which is one of the examples treated by Harnad, Saint-Aubin, and Shnider[12]: their "vacuum" solution is manifestly positive definite, but it is not clear whether all their multisoliton solutions have the same property (they are only given implicitly in Ref. 12!).

ACKNOWLEDGMENTS

It is a pleasure to acknowledge lively discussions about the main points of this work with M. Forger, M. Jacques, and Y. Saint-Aubin. We also thank the referee for pointing out some inaccuracies in an earlier version of the paper.

BP thanks the Institut Interuniversitaire des Sciences Nucléaires (Belgium) for financial support. We both acknowledge the kind hospitality of the Centre de Recherches Mathématiques, Université de Montréal, where part of the work was performed, as well as a travel grant from the Accords Culturels Belgique-Québec (CGRI, Communauté Française de Belgique).

[1]K. Pohlmeyer, "Integrable Hamiltonian systems and interactions through quadratic constraints," Commun. Math. Phys. **46**, 207 (1976).
[2]H. Eichenherr and M. Forger, "On the dual symmetry of the non-linear sigma models," Nucl. Phys. B **155**, 381 (1979).
[3]H. Eichenherr and M. Forger, "More about non-linear sigma models on symmetric spaces," Nucl. Phys. B **164**, 528 (1980); **282**, 745(E) (1987).
[4]M. Forger, "Non-linear sigma models on symmetric spaces," in *Non-linear Partial Differential Operators and Quantization Procedures (Proc. Clausthal 1981)*, *Lecture Notes in Mathematics*, Vol. 1037, edited by S. I. Andersson and H-D. Doebner (Springer, Berlin, 1983), pp. 38–80.
[5]M. Jacques and Y. Saint-Aubin, "Construction of an infinite set of symmetry transformations acting on the solution space of the $\mathbb{C}P^{n-1}$ σ models," in *Fields and Geometry 1986 (Proc. Karpacz)*, edited by A. Jadczyk (World Scientific, Singapore, 1986), pp. 575–580.
[6]M. Jacques and Y. Saint-Aubin, "Infinite-dimensional Lie algebras acting on the solution space of various σ models," J. Math. Phys. **28**, 2463 (1987).
[7]S. Helgason, *Differential Geometry, Lie Groups and Symmetric Spaces* (Academic, New York, 1978).
[8]E. Cartan, "La théorie des groupes finis et continus et l'Analysis Situs," Mém. Sci. Math., Fasc. XLII (Gauthier–Villars, Paris, 1930).
[9]Y. Saint-Aubin, thèse, Université de Montréal, 1982 (unpublished).
[10]Y. Saint-Aubin, "Bäcklund transformations and soliton-type solutions for σ models with values in real Grassmannian spaces," Lett. Math. Phys. **6**, 441 (1982).

[11]J. Harnad, Y. Saint-Aubin, and S. Shnider, "Bäcklund transformations for nonlinear sigma models with values in Riemannian symmetric spaces," Commun. Math. Phys. **92**, 329 (1984).
[12]J. Harnad, Y. Saint-Aubin, and S. Shnider, "The soliton correlation matrix and the reduction problem for integrable systems," Commun. Math. Phys. **93**, 33 (1984).
[13]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).
[14]E. F. Corrigan, D. B. Fairlie, S. Templeton, and P. Goddard, "A Green function for the general self-dual gauge field," Nucl. Phys. B **140**, 31 (1978).
[15]M. Dubois-Violette and Y. Georgelin, "Gauge theory in terms of projection valued fields," Phys. Lett. B **82**, 251 (1979).
[16]R. Sasaki, "Explicit forms of Bäcklund transformations for the Grassmannian and $\mathbb{C}P^{N-1}$ sigma models," J. Math. Phys. **26**, 1786 (1985).
[17]H. J. Borchers and W. D. Garber, "Local theory of solutions for the $O(2k+1)$ σ model," Commun. Math. Phys. **72**, 77 (1980).
[18]W. J. Zakrzewski, "Classical solutions to $\mathbb{C}P^{n-1}$ models and their generalizations," in *Integrable Quantum Field Theories* (*Proc. Tvärminne 1981*), edited by J. Hietarinta and C. Montonen, *Lecture Notes in Physics*, Vol. 151 (Springer, Berlin, 1982).
[19]W. J. Zakrzewski, "Classical solutions of two-dimensional Grassmannian σ models," J. Geom. Phys. **1**, 39 (1984).
[20]R. Sasaki, "Local theory of solutions for the complex Grassmannian and $\mathbb{C}P^{n-1}$ sigma models," Z. Phys. C **24**, 163 (1984).
[21]J. Bognar, *Indefinite Inner Product Spaces* (Springer, Berlin, 1974).
[22]M. Berger, "Les espaces symétriques non compacts," Ann. Sci. Ec. Norm. Sup. **74**, 85 (1957).
[23]R. A. Shapiro, "Pseudo-Hermitian symmetric spaces," Comment. Math. Helv. **46**, 529 (1971).
[24]R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley, New York, 1974).
[25]A. Kundu, "Gauge equivalence of σ-models with non-compact Grassmannian manifolds," J. Phys. A: Math. Gen. **19**, 1303 (1986).
[26]P. Breitenlohner, D. Maison, and G. Gibbons, "4-dimensional black holes from Kaluza–Klein theories," preprint MPI-PAE/PTh 27/87 (München).
[27]J. A. Wolf, *Spaces of Constant Curvature* (Publish or Perish, Boston, 1974), 3rd ed.
[28]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1969), Vol. II.
[29]M. Forger (private communication); Erratum to Ref. 3.
[30]V. E. Zakharov and A. B. Shabat, "Integration of nonlinear equations of mathematical physics by the method of inverse scattering," Funct. Anal. Appl. **13**, 166 (1979).
[31]V. E. Zakharov and A. V. Mikhailov, "On the integrability of classical spinor models in two-dimensional space-time," Commun. Math. Phys. **74**, 21 (1980).

Path-space formulas for covariant gauges in Krein spaces

J. L. Challifour[a] and W. A. Wood
Department of Physics, Indiana University, Bloomington, Indiana 47405

(Received 25 March 1987; accepted for publication 24 June 1987)

Path-space representations in terms of Feynman gauge stochastic integrals are given for the free Maxwell field in all covariant gauges interpolating between the Feynman and Landau gauges. These are applied to quantum electrodynamics and Yang–Mills theories with spatial lattice cutoffs.

I. INTRODUCTION

Present techniques for rigorously constructing quantized theories of gauge fields begin with a regularized theory and then after renormalization study limiting processes by which the regularization is removed. The success of this program depends crucially on nonperturbative *a priori* bounds for the regularized theory and in this respect some regularizations are better than others. Lattice regularizations of the type suggested by Wilson[1] provide a gauge-invariant cutoff theory and have been studied extensively in the Abelian case in Refs. 2–8 and in the non-Abelian case in Refs. 9–12. In Ref. 13, one of us considered a gauge-dependent regularization with a view to obtaining a theory of the gauge potential in an indefinite metric space and thereby a framework closer to that used in theoretical physics; for example, the electroweak interaction; and a Hilbert space which should contain nonzero charge superselection sectors. The inherent difficulty with a non-gauge-invariant regularization is the lack of Osterwalder–Schrader positivity and the stability for the Hamiltonian which it implies. For indefinite metrics, it is shown in Ref. 11 that the Hamiltonian need not be bounded below, depending upon the choice of gauge. This is not surprising but it presents difficulties when studying the removal of cutoffs. For two-dimensional QED, the theory is known to be ultraviolet stable for the gauge potential[7,8] but the construction was carried out in the Landau gauge which is not Osterwalder–Schrader (OS) positive. It is also ultraviolet stable in the Feynman gauge[5] which is OS positive. The obvious question is now raised–how are the two theories related? For the Hamiltonians, the respective operators are sectorial with the sector angle depending upon the choice of gauge. For a cutoff theory, this is proved here and for the case of the continuum limit a proof will be given elsewhere.

The present paper concentrates on obtaining stochastic integral representations for the covariant Rideau gauges[14] for the free Maxwell field in a Krein space. In Sec. II, these gauges are described in terms of the Feynman gauge potential. Explicit gauge transformations interpolating from one gauge to the other are given by Krein unitary operators in Sec. III. We hope these examples will help to focus on a more specific notion of operator gauge transformation than we have found in the literature. In Sec. IV, path-space formulas are given for the relevant semigroups for the interpolating gauges and these are applied in Sec. V to the case of QED and in Sec. VI to Yang–Mills, both with cutoffs. As a by-product, we obtain easy proofs of Krein essential self-adjointness of the Hamiltonians in these cases. This reinforces our view that a complete theory for the gauge potential is now accessible in the Abelian case.

II. KREIN SPACE FOR RIDEAU GAUGES

Throughout this paper, we shall use the formalism of quantum field theory on a finite s-dimensional periodic lattice V with volume $|V|$ and spacing $\delta > 0$. The variables $x = (t,\mathbf{x}), k = (k^0,\mathbf{k})$ have continuum variables t, k^0 and lattice variables $\mathbf{x} = \delta\mathbf{n} \in V$, where n is an integer s-tuple, $\mathbf{k} \in \Gamma_0$ with Γ_0 the lattice dual to V in the Fourier transform

$$f_\mu(\mathbf{x}) = |V|^{-1/2} \sum_{\mathbf{k}\in\Gamma_0} \hat{f}_\mu(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}},$$

with

$$\mathbf{k}\cdot\mathbf{x} = \sum_{j=1}^{s} (\pi n_j\, m_j \delta / L_j), \quad \mu = 0,1,2,\ldots,s,$$

$$|V| = \Pi_{j=1}^{s} (2L_j),$$

while

$$L_2(\Gamma_0) \cong L_{2,0}(V) = \left\{ f \in L_2(V) \,\Big|\, \sum_{x \in \text{two period}} f_\mu(x) = 0 \right\}.$$

On the lattice, we use midpoint approximations to derivatives so either $(\mathbf{k})^j = 2\sin(k^j\delta/2)/\delta$, $\omega(k)^2 = \Sigma_{j=1}^s 4\sin^2(k^j\delta/2)/\delta^2$ or $(\mathbf{k})^j = \sin(k^j\delta)/\delta$, $\omega(k)^2 = \Sigma_{j=1}^s \sin^2(k^j\delta)/\delta^2$. The lattice Maxwell field $A_\mu(k,\mathbf{x})$ is realized on a Fock space H [infinite symmetric tensor product space over $L_{2,0}(V)$] appropriate for an irreducible cyclic representation of the canonical commutation relations

$$[A_\mu(k,\mathbf{x}), \pi_\nu(k,\mathbf{y})] = i g_{\mu\nu}\, \delta_{\mathbf{x},\mathbf{y}} \cdots \tag{2.1}$$

in the Feynman gauge; $g_{00} = -g_{jj} = 1$, $g_{\mu\nu} = 0$ if $\mu \neq \nu$. The Hilbert space H is a Krein space with respect to the Gupta–Bleuler indefinite metric

$$\{\Phi,\Psi\} = (\Phi, \eta\Psi), \quad \eta^* = \eta, \quad \eta^2 = 1,$$

and the annihilation–creation forms a_μ, $a_\mu^\dagger$ satisfy

$$[a_\mu(\mathbf{k}), a_\nu^\dagger(\mathbf{k}')] = -\omega(k) g_{\mu\nu} \delta_{\mathbf{k},\mathbf{k}'}$$

with Fock vacuum $a_\mu(\mathbf{k})\Omega_0 = 0$. Hilbert adjoints are to be denoted by (*) and Krein adjoints by (†). With this notation, the field operator with gauge parameter ξ is given as the formal expression

[a] Also in the Department of Mathematics, Indiana University, Bloomington, Indiana 47405.

 0022-2488/87/112763-12$02.50

$$A_\mu(t,\mathbf{x};\xi) = (2|V|)^{-1/2} \sum_{\mathbf{k}\in\Gamma_0} \frac{1}{\omega(k)} \left[a_\alpha^\dagger(\mathbf{k}) \left(g^{\alpha\beta} + \frac{\xi}{2} \frac{k^\alpha k^\beta}{k_0^2} \right) \left(g_{\beta\mu} e^{ikx} - \frac{\xi k_\beta}{k_0} \frac{\partial}{\partial k_0} \{ k_\mu e^{ik\cdot x} \} \right) \right.$$

$$\left. \left. + a_\alpha(\mathbf{k}) \left(g^{\alpha\beta} + \frac{\xi}{2} \frac{k^\alpha k^\beta}{k_0^2} \right) \left(g_{\beta\mu} e^{-ik\cdot x} - \frac{\xi k_\beta}{k_0} \frac{\partial}{\partial k_0} \{ k_\mu e^{-ik\cdot x} \} \right) \right] \right|_{k_0=\omega}, \tag{2.2}$$

wherein $k\cdot x = k^0 x^0 - \mathbf{k}\cdot\mathbf{x}$. The Feynman gauge arises when $\xi = 0$ and the Landau gauge for $\xi = \frac{1}{2}$. These are the interpolating Rideau gauges that we wish to study. A straightforward argument shows $A_\mu(\xi)$ may be realized as a densely defined, Krein symmetric, linear operator on the subspace D_F of vectors in H with finitely many particles. In fact, vectors in D_F are analytic vectors for A_μ whose closure $\widetilde{A}_\mu = A_\mu^{\dagger\dagger} = A_\mu^{**}$ is then Krein self-adjoint. All relations involving A_μ are understood in this sense as relations between densely defined, closable, unbounded linear operators in H. Whenever possible we shall suppress variables when the context is clear.

The two-point function for (2.2) is given by

$$(\Omega_0, A_\mu(x;\xi) A_\nu(y;\xi)\Omega_0)$$
$$= |V|^{-1} \sum_{\mathbf{k}\in\Gamma_0} \int dk_0\, \theta(k_0)\delta'(k^2)$$
$$\times [g_{\mu\nu}k^2 - 2\xi k_\mu k_\nu] e^{-ik(x-y)},$$

whose Fourier transform has support on the positive energy shell. Time translation is implemented by the Hamiltonian

$$H_0(\xi) = -\sum_{\mathbf{k}\in\Gamma_0} \left. \left(g^{\mu\nu} - 2\xi \frac{k^\mu k^\nu}{\omega^2} \right) \right|_{k_0=\omega} a_\mu^\dagger(\mathbf{k}) a_\nu(\mathbf{k}),$$

which for $0 \leqslant \xi \leqslant 1$ is realized as a sectorial operator with vertex zero and angle $\tan^{-1}(\xi/\sqrt{1-\xi^2})$. The group $e^{itH_0(\xi)}$ is unbounded for $\xi \neq 0$, t real, but is Krein unitary. However, $e^{-tH_0(\xi)}$ for $t \geqslant 0$ is a Krein self-adjoint, C_0 semigroup. On appropriate domains, e.g., D_F, it is true that

$$A_\mu(t,\mathbf{x};\xi) = e^{itH_0(\xi)} A_\mu(\mathbf{x};\xi) e^{-itH_0(\xi)}$$

for the time zero field $A_\mu(\mathbf{x};\xi)$. Canonical momenta[13] for A_μ are given by

$$\pi_j(\mathbf{x};\xi) = -\dot{A}_j(\mathbf{x};\xi),$$
$$\pi_0(\mathbf{x};\xi) = -\dot{A}_0(\mathbf{x};\xi) - 2\xi B(\mathbf{x}),$$
$$B(\mathbf{x}) = (1-2\xi)^{-1} \partial^\alpha A_\alpha(\mathbf{x};\xi),$$

indicating that the singularity at $\xi = \frac{1}{2}$ is canceled by $\partial^\alpha A_\alpha(x;\frac{1}{2}) = 0$. If Feynman gauge fields are indicated by $A_\mu(x)$, $\pi_\mu(x)$ then a little sorting out produces the following expressions for time-zero fields:

$$A_0(\mathbf{x};\xi) = (1-\xi/2)A_0(\mathbf{x}) + (\xi/2)(-\Delta_s)^{-1}\partial_l \pi_l(\mathbf{x}), \tag{2.3a}$$

$$A_j(\mathbf{x};\xi) = [\delta_{jl} + (\xi/2)(-\Delta_s)^{-1}\partial_j\partial_l] A_l(\mathbf{x}) + (\xi/2)(-\Delta_s)^{-1}\partial_j \pi_0(\mathbf{x}), \tag{2.3b}$$

$$\pi_0(\mathbf{x};\xi) = (1+\xi/2)\pi_0(\mathbf{x}) + (\xi/2)\partial_l A_l(\mathbf{x}), \tag{2.3c}$$

$$\pi_j(\mathbf{x};\xi) = [\delta_{jl} - \xi/2(-\Delta_s)^{-1}\partial_j\partial_l]\pi_l(\mathbf{x}) + (\xi/2)\partial_j A_0(\mathbf{x}). \tag{2.3d}$$

Each of these expressions is Krein symmetric on D_F while A_0, π_0 are skew symmetric and A_j, π_j symmetric with respect to the Hilbert metric. By an analytic vector argument using (2.1), the closures of each expression in (2.3) are Krein self-adjoint but Hilbert normal operators. This is the decomposition mentioned in Ref. 13 (p. 328). It is easily demonstrated that the ξ-dependent terms cancel in the time-zero commutators and the relations (2.1) are valid for all values of ξ. Notice further there is no singularity when $\xi = \frac{1}{2}$ in (2.3c).

The relations (2.3) allow representation of all the Rideau gauges in a Krein space by normal operators. Even though Lorentz covariance is restored only in (2.2), the expressions (2.3) are useful for constructing gauge transformations taking $A_\mu(\xi)$ to $A_\mu(\xi')$ and particularly for obtaining a martingale decomposition for their Euclidean (imaginary time) counterparts in terms of Feynman gauge stochastic integrals.

III. GAUGE TRANSFORMATIONS

The choice of a gauge ξ is determined by a mixture of nonphysical fields to be added to the Coulomb gauge. For the Feynman gauge, the lattice cutoff permits the decomposition

$$H = H_L^{(+)} \oplus H_L^{(-)} \oplus H_T$$

corresponding to $\xi = 0$ in

$$A_L^{(\pm)}(\mathbf{x};\xi) = 2^{-1/2}[A_0(\mathbf{x};\xi) \pm i(-\Delta_s)^{-1/2}\partial_l A_l(\mathbf{x};\xi)], \tag{3.1a}$$

$$A_{T_j}(\mathbf{x};\xi) = [\delta_{jl} + (-\Delta_s)^{-1}\partial_j\partial_l] A_l(\mathbf{x};\xi), \tag{3.1b}$$

for which $A_{T_0} = \partial_j A_{T_j} = 0$ with exactly similar relations defining $\pi_L^{(\pm)}$ and π_T. It is easy to verify the commutation relations

$$[A_L^{(\pm)}(\mathbf{x}), \pi_L^{(\pm)'}(\mathbf{y})] = i\{1(\pm)(\pm)'1\}\delta_{\mathbf{x},\mathbf{y}}/2,$$
$$[A_{T_j}(\mathbf{x}), \pi_{T_l}(\mathbf{y})] = -i\{\delta_{jl} + \partial_j\partial_l(-\Delta_s)^{-1}\}\delta_{\mathbf{x},\mathbf{y}};$$

all other commutators vanishing. From (2.3), we learn

$$A_L^{(\pm)}(\mathbf{x};\xi) = (1-\xi/2)A_L^{(\pm)}(\mathbf{x}) \mp i(\xi/2)(-\Delta_s)^{-1/2}\pi_L^{(\pm)}(\mathbf{x}), \tag{3.2a}$$

$$\pi_L^{(\pm)}(\mathbf{x};\xi) = (1+\xi/2)\pi_L^{(\pm)}(\mathbf{x}) \mp i(\xi/2)(-\Delta_s)^{-1/2}A_L^{(\pm)}(\mathbf{x}), \tag{3.2b}$$

$$A_T(\mathbf{x};\xi) = A_T(\mathbf{x}), \quad \pi_T(\mathbf{x};\xi) = \pi_T(\mathbf{x}).$$

These relations show how the choice of gauge parameter ξ adjusts the appropriate mixture of longitudinal modes $A_L^{(\pm)}$ relative to the subspace H_T upon which the physical and Krein metrics agree.

To implement gauge transformations between the unphysical modes consider operators

$$S^{(\pm)} = \pm \frac{\xi}{4} \sum_{x\in V} \delta^s [(-\Delta_s)^{-1/4} \pi_L^{(\pm)}(\mathbf{x}) \mp i(-\Delta_s)^{1/4} A_L^{(\pm)}(\mathbf{x})]^2 \qquad (3.3)$$

defined initially on D_F. With respect to adjoint operations

$$A_L^{(\pm)}(\mathbf{x}) \subset A_L^{(\mp)}(\mathbf{x})^\dagger, \quad \pi_L^{(\pm)}(\mathbf{x}) \subset \pi_L^{(\mp)\dagger}(\mathbf{x}),$$
$$A_L^{(\pm)}(\mathbf{x}) \subset -A_L^{(\pm)}(\mathbf{x})^*, \quad \pi_L^{(\pm)}(\mathbf{x}) \subset -\pi_L^{(\pm)}(\mathbf{x})^*,$$

so that $S^{(\pm)} \subset -S^{(\mp)\dagger}$ but $S^{(\pm)} \not\subset \pm S^{(\pm)*}$. We will now show that on suitable domains $\exp S = \exp\{S^{(+)} + S^{(-)}\}$ is an operator gauge transformation in H which interpolates between the $0 \leqslant \xi \leqslant \frac{1}{2}$ gauges. A proof requires attention to the domains for these unbounded operators.

Consider the operator introduced in Ref. 11,

$$U(\xi) = e^{-i\xi S'} = \exp\Big[-i\xi \sum_{x\in V} \delta^s A_0\, \partial_l A_l(\mathbf{x})\Big]$$
$$= \exp\Big[-\frac{\xi}{2} \sum_{x\in V} \delta^s \{A_L^{(+)}(-\Delta_s)^{1/2} A_L^{(+)}(\mathbf{x}) - A_L^{(-)}(-\Delta_s)^{1/2} A_L^{(-)}(\mathbf{x})\}\Big]$$

and let $\alpha = (\alpha_+, \alpha_-, \alpha_T)$ be a multi-index. Let $\Phi(\alpha)$ denote a vector of the form

$$\Phi(\alpha) = P(A) \exp\Big[\sum_{x\in V} \delta^s \Big\{\frac{\alpha_+}{2} A_L^{(+)}(-\Delta_s)^{1/2} A_L^{(+)}(\mathbf{x}) + \frac{\alpha_-}{2} A_L^{(-)}(-\Delta_s)^{1/2} A_L^{(-)}(\mathbf{x}) - \frac{\alpha_T}{2} A_{T_j}(-\Delta_s)^{1/2} A_{T_j}(\mathbf{x})\Big\}\Big]\Omega_0$$
$$= \text{const}\, P(A) \exp\Big[\sum_{x\in V} \delta^s \Big\{\frac{(\alpha_+ + 1)}{2} A_L^{(+)}(-\Delta_s)^{1/2} A_L^{(+)}(\mathbf{x}) + \frac{(\alpha_- + 1)}{2} A_L^{(-)}(-\Delta_s)^{1/2} A_L^{(-)}(\mathbf{x}) - \frac{(\alpha_T + 1)}{2} A_{T_j}(-\Delta_s)^{1/2} A_{T_j}(\mathbf{x})\Big\}\Big],$$

where $P(A)$ is a polynomial in the time-zero field A_μ. Clearly $U(\xi)\Phi(\alpha) \in H$ if $\alpha_\pm > \pm\xi - 1$, $\alpha_T > -1$.

Definition 3.1: Let $D(\alpha)$ denote all vectors of the form $\Phi(\alpha)$ for a given multi-index α.

Where $\alpha > -1$, $D(\alpha)$ is a dense linear subspace in H consisting of vectors with exponential decrease of order 2 and type $\alpha - 1$. The vectors $D(\alpha)$ are very natural when dealing with $U(\xi)$ as a multiplication operator. It is easily seen at the formal level

$$e^S = U(1) e^M U(-1) \qquad (3.4)$$

for $M = M^{(+)} + M^{(-)}$ and

$$M^{(\pm)} = \pm \frac{\xi}{4} \sum_{x\in V} \delta^s \pi_L^{(\pm)}(-\Delta_s)^{-1/2} \pi_L^{(\pm)}(\mathbf{x}). \qquad (3.5)$$

Lemma 3.2: For any $\alpha > -1$, $D(\alpha)$ consists of entire vectors for $A_L^{(\pm)}$, $\pi_L^{(\pm)}$. If ξ is real and $\alpha > (-\xi - 1, \xi - 1, -1)$ then

$$U(\xi) A_\mu(\mathbf{x}) U(-\xi)\Phi(\alpha) = A_\mu(\mathbf{x})\Phi(\alpha), \qquad (3.6)$$

$$U(\xi)\pi_L^{(\pm)}(\mathbf{x}) U(-\xi)\Phi(\alpha) = \{\pi_L^{(\pm)}(\mathbf{x}) \mp i\xi(-\Delta_s)^{1/2} A_L^{(\pm)}(\mathbf{x})\}\Phi(\alpha). \qquad (3.7)$$

Proof: To show that $\Phi(\alpha)$ is an analytic vector for A_μ estimate $\|A_\mu(\mathbf{x})^n \Phi(\alpha)\|$ and notice that this is equivalent to the harmonic oscillator expression (see the Appendix)

$$\|q^n P(q) e^{-(\alpha+1)q^2/2}\|_2 \leqslant C_0(\epsilon) \|P(q) e^{-(\alpha+1-\epsilon)q^2/2}\|_2$$

as $|q^n e^{-\epsilon q^2/2}| \leqslant (n/\epsilon)^{n/2} e^{-n^2/2} \leqslant C_0(\epsilon)$ uniformly in n. Then $\sum_{n=0}^\infty |\lambda|^n \|A_\mu(\mathbf{x})^n \Phi(\alpha)\|/n!$ converges for all finite λ.

The momentum $\pi_L^{(\pm)}(\mathbf{x})$ is given in harmonic oscillator coordinates as a monomial in differential operators $\partial/\partial q$ for the different modes, so $\pi_L^{(\pm)} D(\alpha) \subset D(\alpha)$. Notice

$$q^k e^{-(\alpha+1)q^2/2} = \frac{1}{\sqrt{2\pi(\alpha+1)}} \int_{-\infty}^{\infty} dp\, e^{ipq} Q(p) e^{-p^2/(2(\alpha+1))}$$

for $Q(p)$ a polynomial of degree k. Under Fourier transform $\Phi(\alpha)$ is mapped into $\Phi(\beta)$ for $\beta = -\alpha/(\alpha+1) > -1$ when $\alpha > -1$. Hence by the Plancherel theorem, vectors in $D(\alpha)$ are entire for $\pi_L^{(\pm)}$ as for $A_L^{(\pm)}$ above.

Relation (3.6) is an identity between multiplication operators, while for (3.7) from the analyticity of the vectors $\Phi(\alpha)$, as $\pi_L^{(\pm)}$ are closable,

$$\pi_L^{(\pm)}(\mathbf{x})U(-\zeta)\Phi(\alpha) = \operatorname*{s-lim}_{N\to\infty}\sum_{n=0}^{N}\pi_L^{(\pm)}(\mathbf{x})\frac{(i\zeta S')^n}{n!}\Phi(\alpha)$$
$$= U(-\zeta)\pi_L^{(\pm)}(\mathbf{x})\Phi(\alpha)$$
$$+\operatorname*{s-lim}_{N\to\infty}\sum_{n=0}^{N}\sum_{m=0}^{n-1}\frac{(i\zeta S')^m}{n!}[\pi_L^{(\pm)}(\mathbf{x}),i\zeta S'](i\zeta S')^{n-1-m}\Phi(\alpha)$$
$$= U(-\zeta)\{\pi_L^{(\pm)}(\mathbf{x})\mp i\zeta(-\Delta_s)^{1/2}A_L^{(\pm)}(\mathbf{x})\}\Phi(\alpha)\,.$$

The right-hand side is clearly in $D(U(\zeta))$ and (3.7) follows. □

Lemma 3.3: The operator e^M is essentially self-adjoint and $e^M D(\alpha)\subset D(\alpha')$ where

$$\alpha'_\pm = \frac{(\alpha_\pm+1)(1\mp\xi/2)-1}{1\pm\xi(\alpha_\pm+1)/2},\quad \alpha'_T=\alpha_T\,.$$

Proof: In terms of harmonic oscillator coordinates,

$$\sum_{x\in V}\delta^s\pi_L^{(\pm)}(\mathbf{x})(-\Delta_s)^{-1/2}\pi_L^{(\pm)}(\mathbf{x}) = \sum_{k\in\Gamma'_0}\omega(k)^{-1}\{p_{1,L}^{(\pm)}(k)^2+p_{2,L}^{(\pm)}(k)^2\},$$

in which $p=-\partial/\partial q$. The result then follows from the calculation

$$e^{(c/2)\partial^2/\partial q^2}P(q)e^{-(\alpha+1)q^2/2} = \frac{Q(q)}{\sqrt{1+c(\alpha+1)}}\exp\left[-\frac{(\alpha+1)}{1+c(\alpha+1)}\frac{q^2}{2}\right],$$

for $c(\alpha+1)>-1$ with Q a polynomial having the same degree as P. □

Proposition 3.4: For $-2<\xi<2$, there exist $\alpha_+>-2$, $\alpha_->0$ such that for $\Phi(\alpha)\in D(\alpha)$

$$e^S A_L^{(\pm)}(\mathbf{x})e^{-S}\Phi(\alpha) = A_L^{(\pm)}(\mathbf{x};\xi)\Phi(\alpha)\,,\quad e^S\pi_L^{(\pm)}(\mathbf{x})e^{-S}\Phi(\alpha)=\pi_L^{(\pm)}(\mathbf{x};\xi)\Phi(\alpha)\,.$$

Proof: For any ξ in the range given, we will choose α so that all operator products are well-defined in the following relations. By Lemma 3.2,

$$A_L^{(\pm)}(\mathbf{x})U(1)e^{-M}U(-1)\Phi(\alpha)$$
$$= U(1)A_L^{(\pm)}(\mathbf{x})e^{-M}\Phi(\alpha_++1,\alpha_--1,\alpha_T)$$
$$= U(1)\operatorname*{s-lim}_{N\to\infty}\sum_{n=0}^{N}\frac{(-1)^n}{n!}A_L^{(\pm)}(\mathbf{x})M^n\Phi(\alpha_++1,\alpha_--1,\alpha_T)$$
$$= U(1)+\operatorname*{s-lim}_{N\to\infty}\sum_{n=0}^{N}\frac{(-M)^n}{n!}A_L^{(\pm)}(\mathbf{x})\Phi(\alpha_++1,\alpha_--1,\alpha_T)$$
$$+U(1)\operatorname*{s-lim}_{N\to\infty}\sum_{n=1}^{N}\frac{(-M)^{n-1}}{(n-1)!}\left\{\mp i\frac{\xi}{2}(-\Delta_s)^{-1/2}\pi_L^{(\pm)}(\mathbf{x})\right\}\Phi(\alpha_++1,\alpha_--1,\alpha_T)$$
$$= U(1)e^{-M}\{A_L^{(\pm)}(\mathbf{x})\mp i(\xi/2)(-\Delta_s)^{-1/2}\pi_L^{(\pm)}(\mathbf{x})\}\Phi(\alpha_++1,\alpha_--1,\alpha_T)\,.$$

Using Lemma 3.3 to choose appropriate α, e^S may be applied to the right-hand side with the result

$$e^S A_L^{(\pm)}(\mathbf{x})e^{-S}\Phi(\alpha) = U(1)\{A_L^{(\pm)}(\mathbf{x})\mp i(\xi/2)(-\Delta_s)^{-1/2}\pi_L^{(\pm)}(\mathbf{x})\}\Phi(\alpha_++1,\alpha_--1,\alpha_T)$$
$$= [A_L^{(\pm)}(\mathbf{x})\mp i(\xi/2)(-\Delta_s)^{-1/2}\{\pi_L^{(\pm)}(\mathbf{x})\mp i(-\Delta_s)^{1/2}A_L^{(\pm)}(\mathbf{x})\}]\Phi(\alpha)$$
$$= A_L^{(\pm)}(\mathbf{x};\xi)\Phi(\alpha)\,.$$

The second to last step uses Lemma 3.2 again. From Lemma 3.3 the indices α should be chosen so that $e^{-S}D(\alpha)\subset D(\alpha')$, where $\alpha'_T=\alpha_T$,

$$\alpha'_+ + 1 = [(\alpha_++2)(1-\xi/2)-1]/[1+(\xi/2)(\alpha_++2)]\,,$$
$$\alpha'_- - 1 = [\alpha_-(1+\xi/2)-1]/[1-\xi\alpha_-/2]\,.$$

A direct calculation verifies that the range $|\xi|<2$ is allowed.

For gauge transformations on the momenta, the same choice of indices α permit

$$U(1)e^M U(-1)\pi_L^{(\pm)}(\mathbf{x})U(1)e^{-M}U(-1)\Phi(\alpha)$$
$$= U(1)e^M[\pi_L^{(\pm)}(\mathbf{x})\pm i(-\Delta_s)^{1/2}A_L^{(\pm)}(\mathbf{x})]e^{-M}U(-1)\Phi(\alpha)$$
$$= U(1)\pi_L^{(\pm)}(\mathbf{x})U(-1)\Phi(\alpha)\pm i(-\Delta_s)^{1/2}A_L^{(\pm)}(\mathbf{x};\xi)\Phi(\alpha)$$
$$= \{\pi_L^{(\pm)}(\mathbf{x})\mp i(-\Delta_s)^{1/2}A_L^{(\pm)}(\mathbf{x})\pm i(-\Delta_s)^{1/2}A_L^{(\pm)}(\mathbf{x};\xi)\}\Phi(\alpha)$$
$$= \pi_L^{(\pm)}(\mathbf{x};\xi)\Phi(\alpha)$$

in which repeated use is made of Lemmas 3.2 and 3.3 and, in the second step, the first part of the proposition. □

Proposition 3.5: The closure of e^S is Krein unitary.

Proof: On suitable $D(\alpha)$, e^S is realized as a product of three essentially Krein unitary operators. For $U(\pm 1)$ this is obvious upon realizing them by multiplication operators. For e^M notice that M is essentially self-adjoint and essentially Krein skew adjoint on allowed $D(\alpha)$. Clearly e^S has a densely defined adjoint so it is closable. The closure must be Krein unitary. □

From Propositions 3.4 and 3.5, there is a Krein unitary equivalence by unbounded operator gauge transformations between the covariant gauges for $|\xi| < 2$. The Krein transformations have a core consisting of analytic vectors for A_μ, π_μ and each gauge transformation leaves pointwise invariant the subspace H_T and thus also pointwise invariant the free Maxwell equations in the Coulomb gauge.

IV. PATH-SPACE FORMULAS

In studying renormalization and other properties of gauge theories, it is a reformulation of the Minkowski quantum field theory in terms of a Euclidean framework that has proved profitable in recent years. The Euclidean framework is usually a symmetric stochastic process and complete details are given for the Feynman gauge in Ref. 13. For the other Rideau gauges, a martingale decomposition for their Euclidean counterparts is not so immediate but is needed in the case of momentum dependent interactions such as appear for the Yang–Mills theory. Equations (2.3) solve this problem in terms of the Feynman gauge fields in that for $\xi \neq 0$, $A_\mu(\mathbf{x};\xi)$ is momentum dependent and correspondingly given by a stochastic integral. It should be noted that the Hamiltonian for $\xi \neq 0$ is not the Feynman gauge Hamiltonian and this requires some care.

Let $B_\mu(t,\mathbf{x};\xi)$ denote a Gaussian process on $R \times V$ with mean zero and covariance

$$\langle B_\mu(t,\mathbf{x};\xi) B_\nu(s,\mathbf{y};\xi)\rangle = \frac{1}{2\pi|V|}\int_{-\infty}^{\infty} dk_0 \times \sum_{\mathbf{k}\in\Gamma_0} e^{ik_0(k-s) + i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} \frac{\widetilde{C}_{\mu\nu}(k)}{k_0^2 + \omega(k)^2}, \tag{4.1}$$

where $\widetilde{C}_{\mu\nu}(k) = \delta_{\mu\nu} - 2\xi k_\mu k_\nu/(k_0^2 + \omega(k)^2)$. Using the distributional identities

$$\int_{-\infty}^{\infty} dk_0 \frac{k_0^2 e^{ik_0(t-s)}}{(k_0^2+\omega^2)^2} = \frac{\pi}{2\omega} e^{-\omega|t-s|}[1 - |t-s|\omega],$$

$$\int_{-\infty}^{\infty} dk_0 \frac{k_0 e^{ik_0(t-s)}}{(k_0^2+\omega^2)^2} = \frac{\pi i(t-s)}{\omega} e^{-\omega|t-s|},$$

$$\int_{-\infty}^{\infty} dk_0 \frac{e^{ik_0(t-s)}}{(k_0^2+\omega^2)^2} = \frac{\pi e^{-\omega|t-s|}}{2\omega^3}[1 + \omega|t-s|];$$

a series of lengthy but straightforward calculations produces the following matrix elements:

$$(\Omega_0, A_\mu(\mathbf{x};\xi) e^{-|t-s|H_0(\xi)} A_\nu(\mathbf{y};\xi)\Omega_0) = (i\delta_{\mu 0} + \delta_{\mu j})(i\delta_{\nu 0} + \delta_{\nu l})\langle B_\mu(t,\mathbf{x};\xi) B_\nu(s,\mathbf{y};\xi)\rangle,$$

$$(\Omega_0, A_\mu(\mathbf{x};\xi) e^{-|t-s|H_0(\xi)} \pi_l(\mathbf{y};\xi)\Omega_0) = (i\delta_{\mu 0} + \delta_{\mu j})(-i)\langle B_\mu(t,\mathbf{x};\xi)\dot{B}_l(s,\mathbf{y};\xi)\rangle,$$

$$(\Omega_0, A_\mu(\mathbf{x};\xi) e^{-|t-s|H_0(\xi)} \pi_0(\mathbf{y};\xi)\Omega_0) = (i\delta_{\mu 0} + \delta_{\mu j})\langle B_\mu(t,\mathbf{y};\xi)\{\dot{B}_0(s,\mathbf{y};\xi) - [2\xi/(2\xi-1)]\partial_\mu B_\mu(s,\mathbf{y};\xi)\}\rangle, \text{ for } \xi \neq \tfrac{1}{2},$$

$$(\Omega_0, \pi_j(\mathbf{x};\xi) e^{-|t-s|H_0(\xi)} \pi_l(\mathbf{y};\xi)\Omega_0) = (-i)^2\langle \dot{B}_j(t,\mathbf{x};\xi)\dot{B}_l(s,\mathbf{y};\xi)\rangle + \delta_{jl}\delta(t-s)\delta_{\mathbf{x},\mathbf{y}},$$

$$(\Omega_0, \pi_j(\mathbf{x};\xi) e^{-|t-s|H_0(\xi)} \pi_0(\mathbf{y};\xi)\Omega_0) = (-i)\langle \dot{B}_j(t,\mathbf{x};\xi)\{\dot{B}_0(s,\mathbf{y};\xi) - [2\xi/(2\xi-1)]\partial_\mu B_\mu(s,\mathbf{y};\xi)\}\rangle, \quad \text{for } \xi \neq \tfrac{1}{2},$$

$$(\Omega_0, \pi_0(\mathbf{x};\xi) e^{-|t-s|H_0(\xi)} \pi_0(\mathbf{y};\xi)\Omega_0) = [1/(2\xi-1)]\delta(t-s)\delta_{\mathbf{x},\mathbf{y}} + \langle\{\dot{B}_0(t,\mathbf{x};\xi) - [2\xi/(2\xi-1)]\partial_\mu B_\mu(t,\mathbf{x};\xi)\} \times \{\dot{B}_0(s,\mathbf{y};\xi) - [2\xi/(2\xi-1)]\,\partial_\nu B_\nu(s,\mathbf{y};\xi)\}\rangle,$$
$$\text{for } \xi \neq \tfrac{1}{2}. \tag{4.2}$$

The covariances for the Landau gauge $\xi = \frac{1}{2}$ are not actually singular after combining terms for $\xi \neq \frac{1}{2}$ and then taking the limit $\xi \to \frac{1}{2}$. If I_t denotes the embedding map

$$I_t\colon H \to \text{Ran}\,(I_t) \subset L_2(S'_R;\mu_\xi)$$

introduced in Eq. 2.9 of Ref. 13, with μ_ξ the measure associated with the covariance (4.1), then

$$I_t^\dagger I_s = e^{-|t-s|H_0(\xi)} = I_s^\dagger I_t \tag{4.3}$$

as $H_0(\xi)$ is Krein self-adjoint, but

$$I_t^* I_s = e^{-|t-s|H_0(\xi)} \neq I_s^* I_t. \tag{4.4}$$

For example,

$$(\Omega_0, A_0(\mathbf{x},\xi) e^{-(t-s)H_0(\xi)^*} A_j(\mathbf{y};\xi)\Omega_0) = -\overline{(\Omega_0, A_j(\mathbf{y};\xi) e^{-(t-s)H_0(\xi)} A_0(\mathbf{x};\xi)\Omega_0)}$$
$$= i\,\overline{\langle B_j(t,\mathbf{y};\xi) B_0(s,\mathbf{x};\xi)\rangle}$$
$$= \frac{i}{2\pi|V|}\int_{-\infty}^{\infty} dp_0 \sum_{\mathbf{p}\in\Gamma_0} e^{ip_0(t-s) + i\mathbf{p}\cdot(\mathbf{x}-\mathbf{y})}\,\widetilde{C}_{0j}(-p_0,\mathbf{p})$$
$$= -(\Omega_0, A_0(\mathbf{x};\xi) e^{-(t-s)H_0(\xi)} A_j(\mathbf{y};\xi)\Omega_0)$$

since for $\xi \neq 0$, $H_0(\xi)^* \neq H_0(\xi)$. In Feynman gauge, H_0 is self-adjoint and the two-point function above vanishes identically.

The list in (4.2) establishes the following correspondence when $\xi \neq 0$:

Minkowski	Euclidean
$I_s A_0(\mathbf{x};\xi) I_s^*$	$iB_0(s,\mathbf{x};\xi)$
$I_s \pi_0(\mathbf{x};\xi) I_s^*$	$\dot{B}_0(s,\mathbf{x};\xi) - [2\xi/(2\xi-1)]\partial_\mu B_\mu(s,\mathbf{x};\xi)$

with a term $(2\xi-1)^{-1}\delta(t-s)\delta_{\mathbf{x},\mathbf{y}}$ for each $\pi_0\pi_0$ contraction, $\xi \neq \frac{1}{2}$

$I_s A_j(\mathbf{x};\xi) I_s^*$	$B_j(s,\mathbf{x};\xi)$
$I_s \pi_j(\mathbf{x};\xi) I_s^*$	$-i\dot{B}_j(s,\mathbf{x};\xi)$

with a term $\delta_{jl}\delta(t-s)\delta_{\mathbf{x},\mathbf{y}}$ for each $\pi_j\pi_l$ contraction.

These correspondences allow finding path-space formulas for $A_\mu(\xi)$ by means of Wick's theorem and the Gaussian process $B_\mu(\xi)$. The canonical quantization in Sec. II for the

Fermi Lagrangian density

$$\mathscr{L}(\xi) = -\tfrac{1}{2}\partial_\mu A_\nu \partial^\mu A^\nu - [\xi/(2\xi-1)](\partial_\mu A^\mu)^2 \qquad (4.5)$$

is faithfully reflected in the Euclidean expressions with the same singularity at the Landau gauge. This singularity is not present in the Hamiltonian, the field operators, or the covariance.

Relations (4.2)–(4.4) represent the semigroup generated by $H_0(\xi)$ as a diffusion given by the Gaussian process with covaraince (4.1). The discussion in Sec. II, however, suggests a representation of this semigroup in terms of the Feynman gauge Gaussian process. From the Lagrangian density (4.5), the Hamiltonian is given by

$$\begin{aligned} H_0(\xi) &= \sum_{\mathbf{x}\in V} \delta^s{:}\Big[\frac{1}{2}\pi_l(\mathbf{x};\xi)^2 + \frac{1}{2}(2\xi-1)\pi_0(\mathbf{x};\xi)^2 \\ &\quad + 2\xi\pi_0\,\partial_l A_l(\mathbf{x};\xi) + \xi(\partial_l A_l(\mathbf{x};\xi))^2 \\ &\quad - \frac{1}{2}(\partial_l A_0(\mathbf{x};\xi))^2 + \frac{1}{2}(\partial_j A_l(\mathbf{x};\xi))^2\Big]{:} \\ &= H_0 + \frac{\xi}{2}\sum_{\mathbf{x}\in V}\delta^s{:}[\{\pi_0(\mathbf{x}) + \partial_l A_l(\mathbf{x})\}^2 \\ &\quad + \{(-\Delta_s)^{-1/2}\partial_l \pi_l(\mathbf{x}) \\ &\quad - (-\Delta_s)^{1/2}A_0(\mathbf{x})\}^2]{:} \end{aligned}$$

where (2.3) have been used to express $H_0(\xi)$ in terms of Feynman gauge operators. To analyze $H_0(\xi)$ it is convenient to use the operators (3.1) for which H_0 becomes a sum of three harmonic oscillators; namely,

$$\begin{aligned} H_0(\xi) &= \frac{1}{2}\sum_{\mathbf{x}\in V}\delta^s[\pi_T(\mathbf{x})^2 + A_T(-\Delta_s)A_T(\mathbf{x}) \\ &\quad - \pi_L^{(\pm)}(\mathbf{x})^2 - A_L^{(\pm)}(-\Delta_s)A_L^{(\pm)}(\mathbf{x}) \\ &\quad + \xi\{\pi_L^{(-)}(\mathbf{x}) + i(-\Delta_s)^{1/2}A_L^{(-)}(\mathbf{x})\} \\ &\quad \times\{\pi_L^{(+)}(\mathbf{x}) - i(-\Delta_s)^{1/2}A_L^{(+)}(\mathbf{x})\}] + \text{const.} \\ &= H_{0,T} + H_{0,L}(\xi)\,. \end{aligned} \qquad (4.6)$$

Now, exploiting once more the harmonic oscillator coordinates in the Appendix, the longitudinal part of $H_0(\xi)$ is unitarily equivalent to a differential operator

$$\begin{aligned} &H_{0,L}(\xi) \\ &= \sum_{\substack{k\in\Gamma_0' \\ j=1,2}}\Big[-\frac{1}{2}\Big\{\frac{\partial^2}{\partial q_{j,L,k}^{(+)2}} + \frac{\partial^2}{\partial q_{j,L,k}^{(-)2}} - 2\xi\frac{\partial^2}{\partial q_{j,L,k}^{(+)}\partial q_{j,L,k}^{(-)}}\Big\} \\ &\quad + \frac{\omega(k)^2}{2}\{q_{j,L,k}^{(+)2} + q_{j,L,k}^{(-)2} - 2\xi q_{j,L,k}^{(+)}q_{j,L,k}^{(-)}\} \\ &\quad + \xi\omega(k)\Big\{q_{j,L,k}^{(-)}\frac{\partial}{\partial q_{j,L,k}^{(+)}} - q_{j,L,k}^{(+)}\frac{\partial}{\partial q_{j,L,k}^{(-)}}\Big\}\Big] \\ &\quad + \text{const.} \end{aligned} \qquad (4.7)$$

In order to obtain a martingale decomposition for $A_\mu(\xi)$ and $\pi_\mu(\xi)$ it is sufficient to examine (4.7) in two dimensions. The general case follows upon taking a direct sum.

Example: Let $x_1 = x_L^{(+)}$, $x_2 = x_L^{(-)}$ for a given j,k and consider the diffusion generated by the differential operator

$$h_{0,L}(\xi) = -\tfrac{1}{2}a_{ij}\partial_i\partial_j - b_i\partial_i + (\omega^2/2)(x,\sigma^2 x)$$

with

$$[a_{ij}] = \begin{pmatrix} 1 & -\xi \\ -\xi & 1\end{pmatrix},$$

$$[b_i] = \beta\begin{pmatrix} x_1 \\ x_2\end{pmatrix},\quad \beta = \xi\omega\begin{pmatrix} 0 & -1 \\ 1 & 0\end{pmatrix},$$

for $-1\leqslant\xi\leqslant 1$, $a=\sigma^2$, where

$$\sigma = \frac{1}{2}\begin{pmatrix} \sqrt{1+\xi}+\sqrt{1-\xi} & -\sqrt{1+\xi}+\sqrt{1-\xi} \\ -\sqrt{1+\xi}+\sqrt{1-\xi} & \sqrt{1+\xi}+\sqrt{1-\xi}\end{pmatrix}.$$

The semigroup generated by $h_{0,L}(\xi)$ is given by

$$\begin{aligned} &(f\Omega_0, e^{-th_{0,L}(\xi)}g\Omega_0) \\ &= (\omega/\pi)\int_{R^2} dx(0)\int d\mu_w \\ &\quad \times\exp\Big[-\frac{\omega}{2}\|x(0)\|^2\Big]\overline{f(x(0))} \\ &\quad \times\exp[-\omega/2\|x(t)\|^2]g(x(t)) \\ &\quad \times\exp\Big[-\frac{\omega^2}{2}\int_0^t ds(x(s),\sigma^2 x(s))\Big] \end{aligned} \qquad (4.8)$$

in which Ω_0 is the Fock vacuum (harmonic oscillator ground state) and $x(t)$ is a solution of the stochastic differential equation

$$dx(t) = \sigma\, db(t) + \beta x(t)dt\,,\quad b(0)=0\,,$$

for $b(t)$ a two-dimensional Brownian motion for the Wiener measure μ_w. In fact,

$$x(t) = e^{\beta t}x(0) + \int_0^t e^{\beta(t-s)}\sigma\, db(s)$$

so by Itô's formula

$$\begin{aligned} &\exp(-\omega/2[\|x(t)\|^2 - \|x(0)\|^2]) \\ &= \exp\Big(-\omega\int_0^t x(s)\cdot dx(s) - 2\omega(1+\xi^2)t\Big). \end{aligned} \qquad (4.9)$$

The expression (4.8) is rearranged by combining (4.9) with a drift transformation conveniently made in two steps. The first step sets $b'(t) = b(t) + \int_0^t \sigma^{-1}\beta x(s)ds$, $dx = \sigma\, db'$, with $b'(t)$ a new Brownian motion. The second drift change is $b''(t) = b'(t) - \int_0^t(\sigma^{-1}\beta - \omega\sigma)x(s)ds$, with $dx = \sigma\, db'' + (\beta - \omega\sigma^2)\,dt$ and $b''(t)$ a Brownian motion with respect to a Wiener measure μ_w''. Equation (4.8) then becomes

$$\begin{aligned} &(f\Omega_0, e^{-th_{0,L}(\xi)}g\Omega_0) \\ &= \frac{\omega}{\pi}\int_{R^2} dx(0)e^{-\omega\|x(0)\|^2} \\ &\quad \times\int d\mu_w''\, e^{-2\omega(1+\xi^2)t}\,\overline{f(x(0))}g(x(t))\,. \end{aligned} \qquad (4.10)$$

The paths are given by

$$x(t) = e^{(\beta-\omega\sigma^2)t}x(0) + \int_0^t e^{(\beta-\omega\sigma^2)(t-s)}\sigma\, db''(s)\,.$$

These paths are readily converted to a harmonic oscillator process $q(t)$ defined in terms of Brownian motion as

$$q(t) = e^{-\omega t}x(0) + \int_0^t e^{-\omega(t-s)}\,db''(s)\,.$$

Finally if μ_0 is the corresponding unique Gauss measure, then

$$(f\Omega_0, e^{-th_{0,L}(\xi)} g\Omega_0)$$
$$= e^{-2\omega(1+\xi^2)t} \int d\mu_0(q)\ \overline{f(q(0))}\ g(x(t)) \qquad (4.11)$$

with paths found from

$$x(t) = q(t) + e^{-\omega t}\begin{pmatrix} 0 & 0 \\ 2\omega\xi t & 0 \end{pmatrix} q(0)$$
$$+ \int_0^t e^{-\omega(t-s)} \left[\begin{pmatrix} 1 & 0 \\ 2\xi\omega(t-s) & 1 \end{pmatrix} \sigma - I\right]$$
$$\times \{dq(s) + \omega q(s) ds\}.$$

The covariance matrix for this process is readily computed with the result $E(x(t)) = 0$ and

$$E(x_j(t)x_j(s)) = e^{-\omega|t-s|}/2\omega, \quad j = 1,2;$$
$$E(x_1(t)x_2(s)) = \xi\,[s - t\wedge s]\ e^{-\omega|t-s|}.$$

The results of this two-dimensional example carry over to the Hamiltonian (4.6), (4.7) by defining "Brownian lattice fields"

$$b_L^{(\pm)}(\mathbf{x},t) = (2|V|)^{-1/2} \sum_{\mathbf{k}\in\Gamma_0} \{\cos(\mathbf{k}\cdot\mathbf{x}) b_{1,L}^{(\pm)}(\mathbf{k},t)$$
$$+ \sin(\mathbf{k}\cdot\mathbf{x}) b_{2,L}^{(\pm)}(\mathbf{k},t)\},$$
$$b_T(\mathbf{x},t) = (2|V|)^{-1/2} \sum_{\mathbf{k}\in\Gamma_0} \{\cos(\mathbf{k}\cdot\mathbf{x}) b_{1,T}(\mathbf{k},t)$$
$$+ \sin(\mathbf{k}\cdot\mathbf{x}) b_{2,T}(\mathbf{k},t)\},$$

wherein all Brownian motions start at zero and are mutually independent for each $\mathbf{k}\in\Gamma_0$. The vector components may be found from

$$b_L^{(\pm)}(\mathbf{x},t)$$
$$= 2^{-1/2}[b_0(\mathbf{x},t) \pm (-\Delta_s)^{-1/2}\partial_l b_l(\mathbf{x},t)]. \qquad (4.12)$$

From the harmonic oscillator expansions, the time-zero Fock space field $A_\mu(\mathbf{x})$ evolves by the semigroup for $H_0(\xi)$ into the operator $\hat{A}_\mu(t,\mathbf{x})$ given by

$$\hat{A}_T(t,\mathbf{x}) = A_T(t,\mathbf{x}) = e^{-(-\Delta_s)^{1/2}t} A_T(\mathbf{x})$$
$$+ \int_0^t e^{-(-\Delta_s)^{1/2}(t-s)}\, db_T(\mathbf{x},s),$$
$$\begin{pmatrix} \hat{A}_0(t,\mathbf{x}) \\ (-\Delta_s)^{-1/2}\partial_l \hat{A}_l(t,\mathbf{x}) \end{pmatrix}$$
$$= e^{(-\Delta_s)^{1/2}\tilde{\beta}t} \begin{pmatrix} A_0(\mathbf{x}) \\ (-\Delta_s)^{-1/2}\partial_l A_l(\mathbf{x}) \end{pmatrix}$$
$$+ \int_0^t e^{(-\Delta_s)^{1/2}\tilde{\beta}(t-s)} \tilde{\sigma} \begin{pmatrix} db_0(\mathbf{x},s) \\ (-\Delta_s)^{-1/2}\partial_l\, db_l(\mathbf{x},s) \end{pmatrix},$$

with matrices

$$\tilde{\beta} = \begin{pmatrix} \xi - 1 & i\xi \\ i\xi & -\xi - 1 \end{pmatrix}, \quad \tilde{\sigma} = \begin{pmatrix} i\sqrt{1-\xi} & 0 \\ 0 & \sqrt{1+\xi} \end{pmatrix}.$$

Clearly when $\xi = 0$, $\hat{A}_\mu$ becomes the Feynman gauge operator A_μ on H. The natural Euclidean field $\hat{B}_\mu$ for the Minkowski–Euclidean correspondence must satisfy

$$\begin{pmatrix} \hat{B}_0(t,\mathbf{x}) \\ (-\Delta_s)^{-1/2}\partial_l \hat{B}_l(t,\mathbf{x}) \\ \hat{B}_T(t,\mathbf{x}) \end{pmatrix}$$
$$= e^{-(-\Delta_s)^{1/2}t} \begin{pmatrix} 1+\xi(-\Delta_s)^{1/2}t & \xi(-\Delta_s)^{1/2}t & 0 \\ -\xi(-\Delta_s)^{1/2}t & 1-\xi(-\Delta_s)^{1/2}t & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} B_0(0,\mathbf{x}) \\ (-\Delta_s)^{-1/2}\partial_l B_l(0,\mathbf{x}) \\ B_T(0,\mathbf{x}) \end{pmatrix}$$
$$+ \int_0^t e^{-(-\Delta_s)^{1/2}(t-s)} \begin{pmatrix} 1+\xi(-\Delta_s)^{1/2}(t-s) & \xi(-\Delta_s)^{1/2}(t-s) & 0 \\ -\xi(-\Delta_s)^{1/2}(t-s) & 1-\xi(-\Delta_s)^{1/2}(t-s) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{1-\xi}\, db_0(\mathbf{x},s) \\ \sqrt{1+\xi}\,(-\Delta_s)^{-1/2}\partial_l\, db_l(\mathbf{x},s) \\ db_T(\mathbf{x},s) \end{pmatrix}. \qquad (4.13)$$

These formulas simplify when expressed in terms of the Brownian fields defined by (4.12) for

$$\hat{B}_L^{(\pm)}(t,\mathbf{x}) = 2^{-1/2}[B_0(t,\mathbf{x}) \pm (-\Delta_s)^{-1/2}\partial_l B_l(t,\mathbf{x})]$$

which satisfy

$$\begin{pmatrix} \hat{B}_L^{(+)}(t,\mathbf{x}) \\ \hat{B}_L^{(-)}(t,\mathbf{x}) \\ \hat{B}_T(t,\mathbf{x}) \end{pmatrix} = e^{-(-\Delta_s)^{1/2}t} \begin{pmatrix} 1 & 0 & 0 \\ 2\xi(-\Delta_s)^{1/2} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} B_L^{(+)}(0,\mathbf{x}) \\ B_L^{(-)}(0,\mathbf{x}) \\ B_T(0,\mathbf{x}) \end{pmatrix}$$

$$+\int_0^t e^{-(-\Delta_s)^{1/2}(t-s)}\begin{pmatrix}1 & 0 & 0\\ 2\xi(-\Delta_s)^{1/2}(t-s) & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\sigma & 0\\ 0 & I\end{pmatrix}\begin{pmatrix}db_L^{(+)}(\mathbf{x},s)\\ db_L^{(-)}(\mathbf{x},s)\\ db_T(\mathbf{x},s)\end{pmatrix}.$$

To write a path-space formula for the semigroup generated by $H_0(\xi)$ in terms of the Feynman gauge Gauss measure μ_0 let $\beta_\mu(\mathbf{x},t)$ be a Brownian field for this measure. By this we mean that for expectations taken with respect to μ_0 (Eq. 4.6 of Ref. 13)

$$\beta_\mu(\mathbf{x},0)=0\,,\quad \langle\beta_\mu(\mathbf{x},t)\rangle=0\,,\quad \langle\beta_\mu(\mathbf{x},t)\,\beta_\nu(\mathbf{y},s)\rangle=\delta_{\mu\nu}\,\delta_{\mathbf{x},\mathbf{y}}\,(t\wedge s)\,,$$

for which the Feynman gauge Euclidean field satisfies

$$B_\mu(t,\mathbf{x})=e^{-(-\Delta_s)^{1/2}t}B_\mu(0,\mathbf{x})+\int_0^t e^{-(-\Delta_s)^{1/2}(t-s)}\,d\beta_\mu(\mathbf{x},s)\,,\quad t>0\,.$$

A short calculation replaces (4.13) by

$$\begin{aligned}\widehat{B}_T(t,\mathbf{x})&=B_T(t,\mathbf{x})\begin{pmatrix}\widehat{B}_0(t,\mathbf{x})\\ (-\Delta_s)^{1/2}\,\partial_l\widehat{B}_l(t,\mathbf{x})\end{pmatrix}\\
&=\begin{pmatrix}B_0(t,\mathbf{x})\\ (-\Delta_s)^{1/2}\,\partial_lB_l(t,\mathbf{x})\end{pmatrix}+\xi t(-\Delta_s)^{1/2}e^{-(-\Delta_s)^{1/2}t}\begin{pmatrix}1 & 1\\ -1 & -1\end{pmatrix}\begin{pmatrix}B_0(0,\mathbf{x})\\ (-\Delta_s)^{-1/2}\,\partial_lB_l(0,\mathbf{x})\end{pmatrix}\\
&\quad+\int_0^t e^{-(-\Delta_s)^{1/2}(t-s)}\begin{pmatrix}\sqrt{1-\xi}\{1+\xi(-\Delta_s)^{1/2}(t-s)\}-1 & \xi\sqrt{1+\xi}(-\Delta_s)^{1/2}(t-s)\\ -\sqrt{1-\xi}\,\xi(-\Delta_s)^{1/2}(t-s) & \sqrt{1+\xi}\{1-\xi(-\Delta_s)^{1/2}(t-s)\}-1\end{pmatrix}\\
&\quad\times\begin{pmatrix}d\beta_0(\mathbf{x},s)\\ (-\Delta_s)^{1/2}\,\partial_l\,d\beta_l(\mathbf{x},s)\end{pmatrix}.\end{aligned}\tag{4.14}$$

When $\xi=0$, the last two terms vanish. For the path-space formula, let $P(A)$,$Q(A)$ be two polynomials in the time-zero Feynman gauge field $A_\mu(\mathbf{x})$, then using (4.14) we find

$$\begin{aligned}&(P(A)\Omega_0,e^{-|t-s|H_0(\xi)}Q(A)\Omega_0)\\&\quad=\int d\mu_0\,\overline{P(i\widehat{B}_0(t,\cdot),\widehat{B}_j(t,\cdot))}Q(i\widehat{B}_0(s,\cdot),\widehat{B}_l(s,\cdot))\end{aligned}\tag{4.15}$$

and the positions of t and s may be reversed on the right-hand side as $\widehat{B}_\mu(t,\mathbf{x})$ is a stationary process. The relations in (4.14) provide a stochastic integral representation for $A_\mu(\xi)$ in terms of a Feynman gauge Brownian field $\beta_\mu(\mathbf{x},t)$. It is not difficult though rather tedious to use (4.15) to reproduce the expressions in (4.2). Many of the terms in (4.15) actually vanish due to the martingale properties of stochastic integrals in (4.13). For these calculations using Wick's theorem it is expedient to use $\pi_\mu(\mathbf{x})\Omega_0=-i(-\Delta_s)^{1/2}A_\mu(\mathbf{x})\Omega_0$.

The restriction of $e^{-tH_0(\xi)}$ to the Coulomb gauge subspace H_T follows immediately from (4.15) as

$$\begin{aligned}&(P(A)\Omega_0,e^{-tH_0(\xi)}Q(A_T)\Omega_0)\\&\quad=\int d\mu_0\,\overline{P(iB_0(0,\cdot),B_j(0,\cdot))}Q(B_T(t,\cdot))\end{aligned}\tag{4.16}$$

and as expected

$$H_0(\xi)|_{H_T}=\frac{1}{2}\sum_{\mathbf{x}\in V}\delta^s{:}[\pi_T(\mathbf{x})^2+A_T(-\Delta_s)A_T(\mathbf{x})]{:}\,.\tag{4.17}$$

By techniques familiar to the boson field theories, (4.17) defines a positive self-adjoint operator and (4.16) leads to Osterwalder–Schrader positivity on H_T.

V. QUANTUM ELECTRODYNAMICS

Let $\psi(\mathbf{x})$, $\psi^\dagger(\mathbf{x})=\psi^*(\mathbf{x})\gamma^0$ denote Dirac spinors on the lattice torus V where γ^0, $i\gamma^j$, $j=1,2,\ldots,s$, are $(s+1)$-dimensional Hermitian matrices satisfying $\{\gamma^\mu,\gamma^\nu\}=2g^{\mu\nu}$. The canonical anticommutation relations are realized in a fermion Fock space H_f by

$$\{\psi_\alpha(\mathbf{x}),\,\psi_\beta(\mathbf{y})\}=0=\{\psi_\alpha^*(\mathbf{x}),\psi_\beta^*(\mathbf{y})\},$$

$$\{\psi_\alpha(\mathbf{x}),\,\psi_\beta^*(\mathbf{y})\}=\delta_{\alpha\beta}\delta_{\mathbf{x},\mathbf{y}}$$

with the consequence $\|\psi_\alpha(\mathbf{x})\|\leqslant 1$, $\|\psi_\alpha^*(\mathbf{x})\|\leqslant 1$. The Hamiltonian operator for QED is

$$H(\xi)=H_0(\xi)+H_0^D+V_{\rm QED}(\xi)=H_0'(\xi)+V_{\rm QED}(\xi)\,,\tag{5.1}$$

where

$$\begin{aligned}H_0^D&=i\sum_{\mathbf{x}\in V}\delta^S{:}\psi^\dagger(\mathbf{x})[-i\boldsymbol{\gamma}\cdot\nabla+m]\psi(\mathbf{x}){:}\,,\\ V_{\rm QED}(\xi)&=e\sum_{\mathbf{x}\in V}\delta^s{:}\psi^\dagger(\mathbf{x})\gamma_\mu\psi(\mathbf{x}){:}\;A^\mu(\mathbf{x};\xi)\,.\end{aligned}\tag{5.2}$$

The decomposition (2.3) shows the momentum dependence of the interaction for $\xi\neq 0$ in the form $V_{\rm QED}=V_{\rm QED}^{\rm electric}+V_{\rm QED}^{\rm magnetic}$ in which

$V_{\rm QED}^{\rm electric}(\xi)$

$$= e \sum_{\mathbf{x}\in V} \delta^s{:}\psi^*(\mathbf{x})\psi(\mathbf{x}){:} \times \left[\left(1-\frac{\xi}{2}\right)A_0(\mathbf{x}) + \frac{\xi}{2}(-\Delta_s)^{-1}\partial_l\pi_l(\mathbf{x})\right], \tag{5.3a}$$

$V_{\rm QED}^{\rm magnetic}(\xi)$

$$= -e \sum_{\mathbf{x}\in V} \delta^s{:}\psi^*(\mathbf{x})\gamma_0\gamma_j\psi(\mathbf{x}){:} \times\left[\left\{\delta_{jl} + \frac{\xi}{2}(-\Delta_s)^{-1}\partial_j\partial_l\right\}A_l(\mathbf{x}) + \frac{\xi}{2}(-\Delta_s)^{-1}\partial_j\pi_0(\mathbf{x})\right]. \tag{5.3b}$$

Each term in the potential is defined on $H_f \otimes D_F$ as a pair of essentially normal operators which commute on a common core of analytic vectors. Even in the Feynman gauge, $V_{\rm QED}^{\rm electric}$ is skew symmetric so the closure of $V_{\rm QED}$ is not self-adjoint. However, $V_{\rm QED}$ is Krein symmetric so it is closable. As the Dirac operators are bounded on the lattice, it is expected that $V_{\rm QED}$ is a "small" perturbation of $H_0(\xi) + H_0^D$, only now these are sectorial rather than self-adjoint operators.

From the representation (4.6) is it easy to see that $H_0(\xi)$ is sectorial. By completing the square for $\xi\pi_L^{(+)}\pi_L^{(-)}$ and $\xi A_L^{(+)}(-\Delta_s)A_L^{(-)}$ then as a bilinear form on D_F

$$\begin{aligned}\mathrm{Re}(\Phi,H_0(\xi)\Phi) &= (\Phi,H_0\Phi) + \xi\sum_{\mathbf{x}\in V}\delta^s(\Phi,\{\pi_L^{(+)}\pi_L^{(-)}(\mathbf{x}) + A_L^{(+)}(-\Delta_s)A_L^{(-)}(\mathbf{x})\}\Phi) + c\|\Phi\|^2\\ &\geqslant (1-|\xi|)(\Phi,H_0\Phi) + c\|\Phi\|^2,\end{aligned}$$

while by the Schwarz inequality and again completing the square

$$\begin{aligned}|\mathrm{Im}(\Phi,H_0(\xi)\Phi)| &\leqslant |\xi|(\Phi,H_{0,L}\Phi)\\ &\leqslant [|\xi|/(1-|\xi|)]\,\mathrm{Re}(\Phi,\{H_0(\xi)-\gamma(\xi)\}\Phi)\end{aligned}$$

when $|\xi|<1$. Thus $\widetilde{H}_0(\xi)|_{D_F}$ is quasi-m-sectorial for $|\xi|<1$ and hence Krein self-adjoint (Lemma 2.2 of Ref. 11). In fact, for the one-particle operator with $0\leqslant\xi\leqslant 1$, the sector is smaller with semiangle $\tan^{-1}\xi/(1-\xi^2)^{1/2}$ (Proposition 2.1 of Ref. 13).

Useful bounds for the interaction in terms of the semigroup generated by $H_0(\xi)$ may be obtained from the Fock annihilation and creation forms c_μ, c_μ^* which satisfy

$$[c_\mu(\mathbf{k}), c_\nu^*(\mathbf{k}')] = \omega(\mathbf{k})\,\delta_{\mu\nu}\delta_{\mathbf{k},\mathbf{k}'}$$

and $a_\mu(\mathbf{k}) = -c_\mu(\mathbf{k})$, $a_\mu^\dagger(\mathbf{k}) = g_{\mu\nu}c_\nu^*(\mathbf{k})$, whereupon $H_0(\xi)$ becomes

$$H_0(\xi) = \sum_{\mathbf{k}\in\Gamma_0} c_\mu^*(\mathbf{k})\left[\delta_{\mu\nu} - \frac{\xi k_\mu k^\nu}{\omega^2}\right]\Bigg|_{k^0=\omega} c_\mu(\mathbf{k}). \tag{5.4}$$

Using a standard representation of the canonical commutation relations (CCR's), such as

$$\begin{aligned}&[c_\mu(\mathbf{k})\Phi]^{(n)}_{\mu_1\mu_2\cdots\mu_n}(k_1,k_2,\ldots,k_n)\\ &\quad = (n+1)^{1/2}\,\Phi^{(n+1)}_{\mu\mu_1\cdots\mu_n}(k,k_1,\ldots,k_n),\end{aligned} \tag{5.5}$$

then on the one-particle space

$$\|e^{-tH_0(\xi)}\phi\|^2 = \sum_{\substack{\mathbf{k}\in\Gamma_0\\ k^0=\omega}} \frac{e^{-2\omega t}}{\omega}\,\overline{\phi_\alpha(k)}\,M^{\alpha\beta}\phi_\beta(k)$$

in which

$$M_{\alpha\beta} = \delta_{\alpha\beta} + (\xi t/\omega)(k_\alpha k^\beta + k^\alpha k_\beta) + 2\xi^2 t^2 k^\alpha k^\beta.$$

The eigenvalues λ for M are $\lambda = 1$ with multiplicity $s-1$ and $\lambda = (\sqrt{1+\xi^2t^2\omega^2} \pm \xi t\omega)^2$. When $|\xi|\leqslant 1$, clearly $e^{-t\omega}(\sqrt{1+\xi^2t^2\omega^2} + |\xi|t\omega)\leqslant 1$ so $e^{-tH_0(\xi)}$ extends to a contraction on H for $t\geqslant 0$.

Proposition 5.1: For $-1<\xi<1$, the operator $\widetilde{H}_0(\xi)|_{D_F}$ generates a Krein self-adjoint, holomorphic, contraction semigroup $e^{-t\widetilde{H}_0(\xi)}$. The operator $c_\mu(\mathbf{k})e^{-tH_0(\xi)}$ defined on D_F extends uniquely to a bounded operator with

$$\|c_\mu(\mathbf{k})e^{-t\widetilde{H}_0(\xi)}\Phi\| \leqslant \mathrm{const}\; t^{-1/2}e^{-t[\omega(k)-\omega_0]}\|\Phi_\mu(k,\cdot)\| \tag{5.6}$$

for $t>0$, $\omega(k)\geqslant\omega_0>0$ for $\mathbf{k}\in\Gamma_0$.

Proof: To show the norm estimate, use (5.4) and (5.5) directly when $\Phi\in D_F$ to show

$$\begin{aligned}\|c_\mu(\mathbf{k})e^{-tH_0(\xi)}\Phi\|^2 &= \sum_{n=0}^{\infty}(n+1)\sum_{\substack{k_j\in\Gamma_0\\ 1\leqslant j\leqslant n}}\frac{\exp(-2t[\omega(k_1)+\cdots+\omega(k_n)])}{\omega(k_1)\cdots\omega(k_n)}\\ &\quad\times\overline{\Phi^{(n+1)}_{\mu\mu_1\cdots\mu_n}(k,k_1,\ldots,k_n)}M^{\mu\nu}(k)M^{\mu_1\nu_1}(k_1)\cdots M^{\mu_n\nu_n}(k_n)\Phi^{(n+1)}_{\nu\nu_1\cdots\nu_n}(k,k_1,\ldots,k_n).\end{aligned}$$

As $k,k_j\in\Gamma_0$, for which the zero modes are removed, then $\omega(k_j)\geqslant\omega_0>0$ and the bilinear form in the summation is dominated by

$$\begin{aligned}&(n+1)e^{-2n\omega_0 t}(\sqrt{1+\xi^2t^2\omega_0^2} + |\xi|t\omega_0)^{2n}\\ &\quad\leqslant e^{2(\omega_0 t-\theta_0)}/2e(\omega_0 t-\theta_0),\end{aligned}$$

with $|\xi|\omega_0 t\sinh\theta_0 \geqslant \theta_0$. The bound in (5.6) now follows with the constant chosen as $[2e(1-|\xi|)]^{-1/2}$ and extends by continuity to all of H. □

The application of Proposition 5.1 to the QED Hamiltonian is almost immediate. If we denote $v_\mu(\mathbf{x}) = {:}\psi^\dagger(\mathbf{x})\gamma_\mu\psi(\mathbf{x}){:}$, a bounded operator on H_f, and its

Fourier transform (see Sec. II) by $\tilde{v}_\mu(\mathbf{k}), \mathbf{k}\in\Gamma_0$, then $V_{\rm QED}$ is a sum of eight operators of the form

$$\sum_{\mathbf{k}\in\Gamma_0} \tilde{v}_\mu(k)\,\rho(\mathbf{k})c_\mu(\mathbf{k})$$

or their adjoints. The functions $\rho(\mathbf{k})$ are bounded. Suppose $\Phi\in H$, $\Psi\in H_f$, then a typical term in $V_{\rm QED}$ satisfies a bound given by (5.6) as

$$\left|\left|\sum_{\mathbf{k}\in\Gamma_0} c_\mu(\mathbf{k})e^{-t\tilde{H}_0(\xi)}\Phi\otimes\rho(k)\tilde{v}_\mu(\mathbf{k})e^{-tH_0^D}\Psi\right|\right|$$

$$\leqslant \text{const}\, t^{-1/2}\sum_{\mathbf{k}\in\Gamma_0}\|\Phi_\mu(k,\cdot)\|\,\|\rho(\mathbf{k})\tilde{v}_\mu(\mathbf{k})e^{-tH_0^D}\Psi\|$$

$$\leqslant \text{const}\, t^{-1/2}\|\Phi\|\Big(\sum_{\mu,k}\|\tilde{v}_\mu(\mathbf{k})\rho(\mathbf{k})e^{-tH_0^D}\Psi\|^2\Big)^{1/2}$$

$$\leqslant \text{const}\, t^{-1/2}\|\Phi\|\Big(\sum_{\mu,\mathbf{x}}\delta^s\|v_\mu(\mathbf{x})e^{-tH_0^D}\Psi\|^2\Big)^{1/2}$$

$$\leqslant \text{const}\, t^{-1/2}\Big(\sum_{\mathbf{x}\in V}\delta^s\|v_\mu(\mathbf{x})\|^2\Big)^{1/2}\|\Phi\otimes\Psi\|\,.$$

In the third step, the vector-valued form of the Plancherel theorem is used. As each $v_\mu(\mathbf{x})$ is a bounded operator on the fermion space, then

$$\|\tilde{V}_{\rm QED}\, e^{-t\tilde{H}_0'(\xi)}\|\leqslant \text{const}\, t^{-1/2},\quad |\xi|<1\,. \tag{5.7}$$

This estimate leads directly to our result for the Hamiltonian (5.1).

Theorem 5.2: For $-1<\xi<1$, $t>0$, $\text{Ran}(e^{-t\tilde{H}_0'(\xi)})\subset D(V_{\rm QED})$ and $\tilde{V}_{\rm QED}$ is a Phillips perturbation of $\tilde{H}_0(\xi)$. Hence $\forall\epsilon>0$, $\exists c(\epsilon)$ such that

$$\|\tilde{V}_{\rm QED}\Phi\|\leqslant\epsilon\|\tilde{H}_0'(\xi)\Phi\|+c(\epsilon)\|\Phi\|\,,\quad \forall\Phi\in D(H_0'(\xi))$$

and $\tilde{H}_0'(\xi)+\tilde{V}_{\rm QED}$ is the generator of a Krein self-adjoint, holomorphic semigroup with the same sector as for $e^{-t\tilde{H}_0'(\xi)}$. Moreover, the semigroup is given by a norm convergent Duhamel series

$$e^{-t(H_0^*(\xi)+V_{\rm QED})} = e^{-t\tilde{H}_0'(\xi)}+\sum_{n=1}^\infty\int_0^t ds_1\int_0^{s_1}ds_2\cdots\int_0^{s_{n-1}}ds_n \times e^{-(t-s_1)\tilde{H}_0'(\xi)}\,\tilde{V}_{\rm QED}\,e^{-(s_1-s_2)\tilde{H}_0'(\xi)} \times\tilde{V}_{\rm QED}\cdots\tilde{V}_{\rm QED}\,e^{-s_n\tilde{H}_0'(\xi)}\,, \tag{5.8}$$

which converges uniformly on compact subsets of the open sector for $e^{-t\tilde{H}_0'(\xi)}$. The Hamiltonian $H_0(\xi)+H_0^D+V_{\rm QED}$ is then essentially Krein self-adjoint.

Proof: Estimate (5.7) establishes the statement about domains and shows $V_{\rm QED}$ to be a Phillips perturbation of $H_0'(\xi)=H_0(\xi)+H_0^D$. Moreover, as $V_{\rm QED}$ has a normal closure and $\tilde{H}_0(\xi)$ is sectorial,

$$\tilde{V}_{\rm QED}R(\lambda,\tilde{H}_0'(\xi)) = \int_0^\infty dt\, e^{-\lambda t}\tilde{V}_{\rm QED}\,e^{-t\tilde{H}_0'(\xi)},\quad \text{Re}\,\lambda>0\,,$$

or

$$\|\tilde{V}_{\rm QED}R(\lambda,H_0'(\xi))\|\leqslant \text{const}\int_0^\infty dt\,\frac{e^{-\text{Re}\,\lambda t}}{t^{1/2}}\leqslant\epsilon$$

when $\text{Re}\,\lambda\geqslant\lambda_0(\epsilon)$. Now if $\Phi\in D(\tilde{H}_0'(\xi))$ then $\Phi=R(\lambda,\tilde{H}_0'(\xi))\Psi$ for some $\Psi\in H\otimes H_f$, hence

$$\|\tilde{V}_{\rm QED}\Phi\|\leqslant\epsilon\|\tilde{H}_0'(\xi)\Phi\|+c(\epsilon)\|\Phi\|\,.$$

The existence of a semigroup for $H(\xi)$ given by a Duhamel series (5.8) is a well known result for Phillips perturbations (Theorem 13.7.2, p. 418 of Ref. 15). The generator is the closure of $H(\xi)|_{D_F}$ which is then essentially quasi-*m*-accretive and so essentially Krein self-adjoint (Lemma 2.2 of Ref. 11). □

From the Duhamel formula (5.8), a path-space formula for the semigroup $e^{-t\tilde{H}(\xi)}$ readily follows in exactly the same manner as for the case of the Yukawa model.[16] There are minor changes due to the γ-matrices; for example,

$$(\Omega_0,e^{-tH(\xi)}\Omega_0)=\int d\mu_\xi\,\det[1+K]\,,$$

$$(\Omega_0,\psi_\alpha(\mathbf{x})e^{-tH(\xi)}\,\psi_\beta^\dagger(\mathbf{y})\Omega_0) = \int d\mu_\xi\,\det[1+K]([1+K]^{-1}S^{\rm E})(t,\mathbf{x};0,\mathbf{y})\,,$$

in which μ_ξ is a Gauss measure for the process $B_\mu(\xi;t,\mathbf{x})$ in (4.1) and $S^{\rm E}$ is the Euclidean fermion propagator

$$S^{\rm E}(x_1-x_2) = (2\pi|V|)^{-1}\int_{-\infty}^\infty dp_0\sum_{\mathbf{p}\in\Gamma_0}e^{ip(x_1-x_2)}\,\frac{(m-i\gamma_\mu^{\rm E}p_\mu)}{p^2+m^2}\,,$$

with Euclidean variables x_1, x_2 and $\gamma_0^{\rm E}=\gamma_0$, $\gamma_j^{\rm E}=-i\gamma^j$ so $\{\gamma_\mu^{\rm E},\gamma_\nu^{\rm E}\}=2\delta_{\mu\nu}$. The operator K has matrix elements

$$K(x_1,x_2) = ie_0S^{\rm E}(x_1-x_2)\gamma_\mu^{\rm E}\,B_\mu(\xi;x_2)\chi_{[0,t]}(1-\delta_{x_1,x_2})$$

for $x_1=(t,\mathbf{x})$, $x_2=(0,\mathbf{y})$ and $\text{tr}\,\gamma_\mu^{\rm E}=0$ implies $\text{tr}\,K=0$. The diagonal terms for K are removed by the Wick ordering in $V_{\rm QED}$.

On restricting $H(\xi)$ to the Coulomb gauge subspace $H_T\otimes H_f$ as defined in (3.1) and (3.2), then

$$H(\xi)|_{\text{Coulomb}\cap D_F} = H_{0,T}+H_0^D-e\sum_{\mathbf{x}\in V}\delta^s{:}\,\psi^\dagger(\mathbf{x})\,\gamma_j\psi(\mathbf{x}){:}\,A_{T,j}(\mathbf{x})\,.$$

This is essentially self-adjoint and bounded below with closure the generator of a self-adjoint contraction semigroup on $H_T\otimes H_f$. The path-space formula in this case now possesses Osterwalder–Schrader positivity.

VI. ESSENTIAL SELF-ADJOINTNESS FOR YANG–MILLS

To show $(H_0(\xi)+V_{\rm YM}^{(\xi)})|_{D_F}$ is essentially Krein self-adjoint for $\xi\neq0$, we exploit the gauge transformation of Sec. III. The Yang–Mills interaction on the lattice V for a general covariant gauge is given (Eq. 3.3 of Ref. 13) as a sum of two operators

$$V_{\rm YM}^{\rm magnetic}(\xi)=\sum_{\mathbf{x}\in V}\delta^s\Big[\lambda\,\partial_jA_l\cdot A_j\times A_l(\mathbf{x};\xi) +\frac{\lambda^2}{4}(A_j\times A_l(\mathbf{x};\xi))^2\Big]\,, \tag{6.1}$$

$$V_{\rm YM}^{\rm electric}(\xi) = \lambda \sum_{\mathbf{x}\in V} \delta^s [\pi_l(\mathbf{x};\xi) + \partial_l A_0(\mathbf{x};\xi)] \cdot A_0 \times A_l(\mathbf{x};\xi), \quad (6.2)$$

in which sums over color indices are suppressed. In Feynman gauge denote $V_{\rm YM}(0) = V_{\rm YM}$ and consider the operator

$$H''_{\rm YM} = H_0 + \xi V_0 + V_{\rm YM} \quad (6.3)$$

for

$$V_0 = \sum_{\mathbf{x}\in V} \delta^s :\{\pi_0(\mathbf{x}) + \partial_l A_l(\mathbf{x})\}^2: . \quad (6.4)$$

The operator in (6.3) is readily seen to be defined on the domains $D(\alpha)$ of Definition 3.1. From Proposition 3.4 for a suitable choice of α

$$H_{\rm YM} = H_0(\xi) + V_{\rm YM}(\xi) = e^S H''_{\rm YM} e^{-S}$$

as the expressions (6.1), (6.2), and (6.4) do not involve ξ explicitly. Recall $e^S = U(1)e^M U(-1)$, one finds further that

$$H_{\rm YM} = U(1)e^M H'''_{\rm YM}\, e^{-M} U(-1),$$

where [Eqs. (2.5), (2.6), (2.7) of Ref. 11]

$$H'''_{\rm YM} = H'_{\rm YM} + \xi \sum_{\mathbf{x}\in V} \delta^s :\pi_0(\mathbf{x})^2:$$
$$= \sum_{\mathbf{x}\in V} \delta^s \Big[\frac{\pi_l(\mathbf{x})^2}{2} - \frac{(1-2\xi)\pi_0(\mathbf{x})^2}{2} + \frac{F_{jl}(\mathbf{x})^2}{4} + \lambda \pi_l \cdot A_0 \times A_l(\mathbf{x}) + \pi_0\, \partial_l A_l(\mathbf{x}) - \pi_l\, \partial_l A_0(\mathbf{x}) \Big] - E_0(\xi) \quad (6.5)$$

with

$$E_0(\xi) = \frac{s+1-\xi}{2} \sum_{\mathbf{p}\in\Gamma_0} \omega(p).$$

By transforming to harmonic oscillator coordinates $H'''_{\rm YM}$ is realized as unitarily equivalent to an elliptic operator of the form (Eq. 2.8 of Ref. 11)

$$H'''_{\rm YM} = -(a_{ij}/2)\partial_i \partial_j + \mathbf{a}\cdot\nabla + V(q),$$

in which

$$[a_{ij}] = \left[\begin{pmatrix} I & 0 \\ 0 & (1-2\xi)I \end{pmatrix}\right].$$

When $\xi \leqslant \frac{1}{2}$, the elliptic operator in (6.5) is only degenerate at the Landau gauge. However, even in this case the proof that $\widetilde{H}'''_{\rm YM}|_{C_0^\infty}$ is quasi-m-accretive given in Theorem 2.4 of Ref. 11 is unchanged. Consequently, $\widetilde{H}'''_{\rm YM}|_{D_F}$ is Krein self-adjoint and $\widetilde{H}_{\rm YM}(\xi)$ is Krein self-adjoint by Proposition 3.5. We summarize this discussion in the following theorem.

Theorem 6.1: The Yang–Mills Hamiltonian $H_{\rm YM}(\xi) = H_0(\xi) + V_{\rm YM}(\xi)$ for $-2 < \xi \leqslant \frac{1}{2}$ is essentially Krein self-adjoint on $D(\alpha)$, for α chosen as in Proposition 3.4.

Remark: While this last result allows definition of a unique Yang–Mills theory with cutoffs which interpolates between the Feynman and Landau gauges, the difficulty that $H''_{\rm YM}$ is bounded below and not $H_{\rm YM}$ is still present in the Rideau gauges. The root of the trouble is that $\Omega_0 \notin D(U(\pm 1))$. Restricting $H_{\rm YM}(\xi)$ to the axial gauge subspace, $A_0(\mathbf{x};\xi) = 0$, does in fact lead to an essentially self-adjoint operator which is bounded below. The same is true for restriction to the Coulomb subspace. Both restrictions forego covariance and locality which is the *raison d'etre* for the indefinite metric.

APPENDIX: HARMONIC OSCILLATOR COORDINATES

By means of Fock annihilation and creation forms $c_\mu(k), c^*_\mu(k)$ for which

$$[c_\mu(k), c^*_\nu(k')] = \omega(k)\delta_{\mu\nu}\, \delta_{k,k'}$$

a representation of (2.1) at time zero is provided by

$$A_\mu(\mathbf{x}) = (2|V|)^{-1/2} \sum_{k\in\Gamma_0} \frac{\{g_{\mu\nu} c^*_\nu(k) e^{-i\mathbf{k}\cdot\mathbf{x}} - c_\mu(k) e^{i\mathbf{k}\cdot\mathbf{x}}\}}{\omega(k)},$$

$$\pi_\mu(\mathbf{x}) = -i(2|V|)^{-1/2} \times \sum_{k\in\Gamma_0} \{g_{\mu\nu} c^*_\nu(k) e^{-i\mathbf{k}\cdot\mathbf{x}} + c_\mu(k) e^{i\mathbf{k}\cdot\mathbf{x}}\}.$$

The q-coordinates are defined by

$$A_\mu(\mathbf{x}) = (2|V|)^{-1/2} \times \sum_{k\in\Gamma'_0} \{q_{1,\mu}(k)\cos(\mathbf{k}\cdot\mathbf{x}) + q_{2,\mu}(k)\sin(\mathbf{k}\cdot\mathbf{x})\},$$

$$\pi_\mu(\mathbf{x}) = (2|V|)^{-1/2} \times \sum_{k\in\Gamma'_0} \{p_{1,\mu}(k)\cos(\mathbf{k}\cdot\mathbf{x}) + p_{2,\mu}(k)\sin(\mathbf{k}\cdot\mathbf{x})\},$$

where Γ'_0 indicates those momenta in the dual lattice remaining after using the relations

$$q_{j,\mu}(k) = (-1)^{j+1} q_{j,\mu}(-k),$$
$$p_{j,\mu}(k) = (-1)^{j+1} p_{j,\mu}(-k).$$

For these momenta, (2.1) requires $[q_{j,\mu}(k), p_{l,\nu}(k')] = i\delta_{jl}\, g_{\mu\nu}\, \delta_{k,k'}$. Each $q_{j,l}$ and $p_{j,l}$ are symmetric but in the Feynman gauge $q_{j,0}$ and $p_{j,0}$ are skew symmetric. An irreducible representation for these Heisenberg relations is provided by $q_{j,l}(k)$ and $-iq_{j,0}(k)$ as multiplication operators with $p_{j,l} = i\,\partial/\partial q_{j,l}$ and $p_{j,0} = -\partial/\partial q_{j,0}$. For the Stokes fields in (3.1) and (3.2), provided one uses a symmetric or midpoint approximation to the lattice derivatives, the oscillator variables are natural. For example, with midpoint derivatives

$$q^{(\pm)}_{1,L}(k) = 2^{-1/2}\{q_{1,0}(k) \pm 2iq_{2,l}(k) \times \sin(k_l\delta/2)/[\delta\omega(k)]\},$$

$$q^{(\pm)}_{2,L}(k) = 2^{-1/2}\{q_{2,0}(k) \mp 2iq_{1,l}(k) \times \sin(k_l\delta/2)/[\delta\omega(k)]\},$$

with analogous expression for $p^{(\pm)}_{j,L}(k)$. Again one finds

$$[q^{(\pm)}_{j,L}(k), p^{(\pm)'}_{l,L}(k')] = (1(\pm)(\mp)'1)\delta_{j,l}\, \delta_{k,k'}/2.$$

Systematic use of these expressions and their Fourier series for $A_\mu(\mathbf{x})\pi_\mu(\mathbf{x})$, $A^{(\pm)}_L(\mathbf{x})$, $\pi^{(\pm)}_L(\mathbf{x})$ lead to the formulas appearing in Lemmas 3.2, 3.3 and Eqs. (4.7) and (4.12).

[1]K. G. Wilson, Phys. Rev. D **10**, 2445 (1974).
[2]T. Balaban, J. Imbrie, and A. Jaffe, *Progress in Gauge Theories*, 1983 Cargèse Lectures, edited by G. Lehmann, G. t'Hooft, A. Jaffe, P. Mitter, I. Singer, and R. Stora (Plenum, New York, 1984), pp. 79–104.
[3]T. Balaban, J. Imbrie, and A. Jaffe, Commun. Math. Phys. **97**, 299 (1985).
[4]D. Brydges, J. Fröhlich, and E. Seiler, Ann. Phys. (NY) **121**, 227 (1979); Commun. Math. Phys. **71**, 159 (1980); **79**, 353 (1981).
[5]J. L. Challifour, unpublished.
[6]L. Gross, Commun. Math. Phys. **92**, 137 (1983).
[7]D. Weingarten and J. L. Challifour, Ann. Phys. (NY) **123**, 61 (1979).
[8]D. Weingarten, Ann. Phys. (NY) **126**, 154 (1980).
[9]T. Balaban, "Renormalization group approach to lattice gauge field theories I. Generation of effective actions in a small field approximation and a coupling constant renormalization in four dimensions," Harvard University preprint HUTMP/B189, 1986.
[10]T. Balaban and A. Jaffe, "Constructive gauge theory," Harvard University preprint HUTMP86/B191, 1986.
[11]J. L. Challifour, Ann. Inst. H. Poincaré **42**, 1 (1985).
[12]P. Federbush, "A phase cell approach to Yang–Mills theory I–VI," University of Michigan preprints.
[13]J. L. Challifour, Ann. Phys. (NY) **136**, 317 (1981).
[14]G. Rideau, Lett. Math. Phys. **1**, 17 (1975); J. Math. Phys. **19**, 1627 (1978).
[15]E. Hille and R. S. Phillips, *Functional Analysis and Semigroups*, rev. ed. (American Math. Soc., Providence, RI, 1957).
[16]E. Seiler and B. Simon, Ann. Phys. (NY) **97**, 470 (1976).

Nonpolynomial Yang–Mills local cohomology

Giuseppe Bandelloni
Istituto di Fisica della Facoltá di Ingegneria-Universitá di Genova, Istituto Nazionale di Fisica Nucleare, Sezione di Genova, Italy

(Received 10 February 1987; accepted for publication 15 July 1987)

The spectral sequence method is used to find the nonintegrated function cohomology of the BRS Yang–Mills operator in the space of the analytic (*a priori* nonpolynomial) functions in space-time dimensions.

I. INTRODUCTION

Symmetries are, in several cases, part of the definition of physical models; they indeed imply conservation laws that restrict the class of possible processes, thus simplifying the task of constructing dynamical models.

This is akin to the widespread use of group theory and differential geometry in classical and quantum mechanics, to simplify the mathematical formulation and to characterize anomalous behavior due to quantum effects.

In particular the BRS[1] perturbative quantization procedure emphasized the role of the cohomology space of the differential operator induced by the symmetry group.

We shall find in this paper the cohomology space of the BRS operator of the pure Yang–Mills model[2]; that is, we solve the equations

$$\boldsymbol{\delta}\Delta(x)=0, \tag{1.1a}$$

$$\Delta(x)\neq\boldsymbol{\delta}\Delta(x), \tag{1.1b}$$

where $\Delta(x)$ is an analytic function, $\boldsymbol{\delta}$ is the Yang–Mills-BRS operator, and $\Delta(x)$ is an arbitrary function.

Our result is

$$\Delta(x)=F(\mathscr{D}^{a*}{}_{\alpha(n)}G^{*}{}_{\mu\nu}(x))T(C^a(x)), \tag{1.2}$$

where $F(\mathscr{D}^{a*}{}_{\alpha(n)}G^{*}{}_{\mu\nu}(x))$ and $T(C^a(x))$ are group global invariant analytic functions, which depend on covariant derivatives of the curvature tensor $G^a{}_{\mu\nu}(x)$, and underivated Faddeev–Popov ghosts, respectively.

It is important to note that we get rid of the polynomiality hypothesis and our result holds for any space-time dimensions.

II. THE MODEL AND THE STRATEGY

The pure Yang–Mills[1] models are based on a vector connection gauge field $A^a{}_\mu(x)$ in a d-dimensional space-time, which carries the adjoint representation of a semisimple local Lie group $\mathscr{G}$.

With the aid of the anticommuting Faddeev–Popov ($\Phi\Pi$) charged fields $C^a(x)$ the local transformations of $\mathscr{G}$ induce the null-squared BRS[1] differential operator

$$\boldsymbol{\delta}=\int d^dx\left\{[\partial_\mu C^a(x)+f^{abc}A^b{}_\mu(x)C^c(x)]\frac{\delta}{\delta A^a{}_\mu(x)}-\frac{1}{2}[f^{abc}C^b(x)C^c(x)]\frac{\delta}{\delta C^a(x)}\right\}, \tag{2.1}$$

whose functional cohomology controls the perturbative renormalization program of the model.

We shall study, in this paper, the cohomology space $H(\boldsymbol{\delta})(x)$ of the operator $\boldsymbol{\delta}$ in the space of the nonintegrated functions; that is we shall solve the system

$$\boldsymbol{\delta}\Delta(x)=0, \tag{2.2a}$$

$$\Delta(x)\neq\boldsymbol{\delta}\Delta(x), \tag{2.2b}$$

for any $\Delta(x)$.

The space of the local functions $\Delta(x)$ can be parametrized[3] by the monomials of the fields and their space-time derivatives, which are to be considered as independent coordinates of the model (since partial integrations are not admissible) and the operator $\boldsymbol{\delta}$, when acting on $\mathscr{S}$, can be written as

$$\boldsymbol{\delta}=\sum_{n=0}\left\{D_{\alpha(n)}[\partial_\mu C^a(x)+f^{abc}A^b{}_\mu(x)C^c(x)]\times\frac{\partial}{\partial D_{\alpha(n)}A^a{}_\mu(x)}-\frac{1}{2}D_{\alpha(n)}[f^{abc}C^b(x)C^c(x)]\frac{\partial}{\partial D_{\alpha(n)}C^a(x)}\right\}, \tag{2.3}$$

where

$$D_{\alpha(n)}A^a{}_\mu(x)=\partial_{\alpha(1)}\partial_{\alpha(2)}\cdots\partial_{\alpha(n)}A^a{}_\mu(x). \tag{2.4}$$

To accomplish our program, we shall use the spectral sequence[4] method, already introduced by Dixon,[3] and used by the author in different cases,[5] which allows one to reach, by iterations, a space $\underline{H}(\boldsymbol{\delta})(x)$ isomorphic to $H(\boldsymbol{\delta})(x)$.

The first ingredient for our recipe is to endow the space $\mathscr{S}$ with a grading induced by an operator ν,

$$\mathscr{S}=\oplus_{k\in Z}\mathscr{S}_k, \tag{2.5}$$

such that each function $f_k(x)\in\mathscr{S}_k$ is eigenfunction of ν with eigenvalue equal to k, and there exists an adjointness relationship with respect to which the operator ν is self-adjoint,

$$\nu=[\nu]^+. \tag{2.6}$$

The action of $\boldsymbol{\delta}$ on $\mathscr{S}_k$ will increase the grading degree, while that of $[\boldsymbol{\delta}]^+$ will lower it,

$$\boldsymbol{\delta}:\ \mathscr{S}_k\Rightarrow\mathscr{S}_{k+r}, \tag{2.7a}$$

$$[\boldsymbol{\delta}]^+:\ \mathscr{S}_k\Rightarrow\mathscr{S}_{k-r}\quad(r\geqslant1). \tag{2.7b}$$

So elements belonging to spaces with different ν eigenvalues will be orthogonal with respect to the scalar product defined by our adjointness procedure.

The BRS operator $\boldsymbol{\delta}$ will be decomposed by ν as

$$[\boldsymbol{\delta},\nu]=\sum_{k=0}k\boldsymbol{\delta}(k). \tag{2.8}$$

 0022-2488/87/112775-06$02.50

The general theorems of Appendix A, in which a more rigorous introduction to the spectral sequences method can be found, show that $\underline{H}(\boldsymbol{\delta})(x)$ is the space of functions $\Delta(x)$ which are solutions of the (*a priori*) infinite system

$$\boldsymbol{\delta}(k)\Delta(x)=0, \tag{2.9a}$$

$$[\boldsymbol{\delta}]^+(k)\Delta(x)=0 \quad (k=1,2,3...). \tag{2.9b}$$

It is obvious that the isomorphism which relates $\underline{H}(\boldsymbol{\delta})(x)$ to $H(\boldsymbol{\delta})(x)$ depends on the choice of the grading operator ν, one might expect to get a neater result if the grading operation would preserve the superselection rules of the model.

Furthermore, if we want to get local functions with definite Faddeev–Popov ($\Phi\Pi$) charge, and to recover the $\Phi\Pi$ charge additivity property, we have to assume that each function $F(x)$ of the space $\mathscr{S}$ be analytical in the field $C^a(x)$ and its space-time derivatives; namely,

$$F(x)=F^0[A^a{}_\mu(x)]+\sum_{n=0}D_{\alpha(n)}C^a(x)F^1{}_a{}^{\alpha(n)}[A^d{}_\mu(x)]$$
$$+\sum_{n=0}\sum_{m=0}D_{\alpha(n)}C^a(x)D_{\beta(m)}C^b(x)$$
$$\times F^2{}_a{}^{\alpha(n)}{}_b{}^{\beta(m)}[A^d{}_\mu(x)], \tag{2.10}$$

where the functions

$$F^0[A^a{}_\mu(x)],\ F^1{}_a{}^{\alpha(n)}[A^d{}_\mu(x)],$$
$$F^2{}_a{}^{\alpha(n)}{}_b{}^{\beta(m)}[A^d{}_\mu(x)], \tag{2.11}$$

are analytic functions on the gauge field $A^d{}_\mu(x)$ and its space-time derivatives for the reasons we shall later specify.

This assumption will imply that the counting operator

$$\nu=\sum_{n=0}(1+n)D_{\alpha(n)}C^a(x)\frac{\partial}{\partial D^{\alpha(n)}\,C^a(x)} \tag{2.12}$$

induces a grading on $\mathscr{S}$, whose degree k is provided by the value of the $\Phi\Pi$ charge plus the number of the derivatives of the $C^a(x)$ fields of the function in each sector $\mathscr{S}_k$.

It is now useful, for later uses, to introduce a new coordinate system in the uncharged $\Phi\Pi$ sector, such that each function of the gauge field $A^a{}_\nu(x)$ and its space-time derivatives can be reparametrized in a different way.

Each derivative of order n of the field $A^c{}_\mu(x)$ can be expressed in terms of its symmetrized and antisymmetrized components

$$D^{\alpha(n)}A^c{}_\mu(x)=D_{\{\alpha(n)}A^c{}_{\mu\}}(x)+D_{\{\alpha(n)}A^c{}_{\mu\}}(x). \tag{2.13}$$

If we introduce the usual curvature tensor

$$G^a{}_{\mu\nu}(x)=\partial_\mu A^a{}_\nu(x)-\partial_\nu A^a{}_\mu(x)$$
$$+f^{abc}A^b{}_\mu(x)A^c{}_\nu(x), \tag{2.14}$$

and the covariant derivative operator

$$\mathscr{D}^{ab}{}_\mu A^b{}_\sigma(x)\equiv[\partial_\mu\delta^{ab}+f^{abc}A^b{}_\mu(x)]A^c{}_\sigma(x), \tag{2.15}$$

it is easy to realize that the antisymmetric part $D_{\{\alpha(n)}A^c{}_{\nu\}}(x)$ can be expressed in terms of covariant derivatives of the curvature tensor and symmetrized derivatives of the gauge field of lower orders,

$$D_{\{\alpha(n-1)}\,\partial_\nu A^c{}_{\mu\}}(x)$$
$$=\mathscr{D}^{c^*}{}_{\alpha(n-1)}G^*{}_{\mu\nu}(x)$$
$$+\mathscr{P}^c{}_{\alpha(n-1),\mu\nu}(\mathscr{D}^{d^*}{}_{\alpha(r)}G^*{}_{\rho\sigma}(x),D_{\{\alpha(s)}A^b{}_{\gamma\}}(x)),$$
$$0\leqslant r<n-2,\quad 0\leqslant s<n-1, \tag{2.16}$$

where we have indicated

$$\mathscr{D}^{a^*}{}_{\alpha(n)}G^*{}_{\mu\nu}(x)$$
$$=\mathscr{D}^{ab(1)}{}_{\alpha(1)}\mathscr{D}^{b(1)b(2)}{}_{\alpha(2)}\cdots\mathscr{D}^{b(n-1)b(n)}{}_{\alpha(n)}$$
$$\times G^{b(n)}{}_{\mu\nu}(x). \tag{2.17}$$

Therefore each space-time derivative of order n of the gauge field $A^b{}_\mu(x)$ can be written in terms of the symmetrized derivative $D_{\{\alpha(k)}A^c{}_{\mu\}}(x)(k\leqslant n)$, and covariant derivative of $G^a{}_{\mu\nu}(x)$ of order less than n.

This implies a one-to-one map between the coordinate system defined by the field $A^c{}_\mu(x)$ and its space-time derivatives on one side, and the one expressed by the *symmetrized* derivative $D_{\{\alpha(k)}A^c{}_{\mu\}}(x)$ $(k=0,1,2,...)$, the curvature tensor $G^a{}_{\mu\nu}(x)$, and its covariant derivatives, on the other side. Thus we are free to choose the latter one for our purposes. With this choice of variables, the BRS operator $\boldsymbol{\delta}$ can be written

$$\boldsymbol{\delta}=\sum_{n=0}\Big\{[D_{\{\alpha(n)}\,\partial_{\mu\}}C^a(x)+f^{abc}D_{\{\alpha(n)}\,[A^b{}_{\mu\}}(x)C^c(x)]]$$
$$\times\frac{\partial}{\partial D_{\{\alpha(n)}A^a{}_{\mu\}}(x)}$$
$$+f^{abc}C^b(x)\mathscr{D}^{c^*}{}_{\alpha(n)}G^*{}_{\mu\nu}(x)\frac{\partial}{\partial\mathscr{D}^{a^*}{}_{\alpha(n)}G^*{}_{\mu\nu}(x)}$$
$$-\frac{1}{2}D_{\alpha(n)}\,[f^{abc}C^b(x)C^c(x)]\frac{\partial}{\partial D_{\alpha(n)}C^a(x)}\Big\}. \tag{2.18}$$

The operator ν decomposes the BRS operator $\boldsymbol{\delta}$ as

$$\boldsymbol{\delta}(1)=\sum_{n=0}\Big\{C^c(x)\,[f^{abc}D_{\{\alpha(n)}A^b{}_{\mu\}}(x)]\,\frac{\partial}{\partial D_{\{\alpha(n)}A^a{}_{\mu\}}(x)}$$
$$-\frac{1}{2}D_{\alpha(n)}\,[f^{abc}C^b(x)C^c(x)]\frac{\partial}{\partial D_{\alpha(n)}C^a(x)}\Big\} \tag{2.19}$$

and

$$\boldsymbol{\delta}(k)=D_{\alpha(k-2)}\,\partial_\mu C^a(x)\Big\{\frac{\partial}{\partial D_{\{\alpha(k-2)}A^a{}_{\mu\}}(x)}$$
$$+\sum_{n=0}\frac{(n+s-1)!}{n!(s-1)!}f^{cba}D_{\{\alpha(n)}A^b{}_{\nu\}}(x)$$
$$\times\frac{\partial}{\partial D_{\{\alpha(n)}\,\partial_\mu A^c{}_{\nu\}}(x)}\Big\}$$
$$\equiv D_{\alpha(k-2)}\,\partial_\mu C^a(x)\mathscr{N}^a{}_{\alpha(k-2),\mu}(x),$$
$$k=2,3,..., \tag{2.20}$$

in which the curvature tensor content disappears.

The adjointness relationship is defined, as in Ref. 3, by the trivial substitution of each monomial in the fields with the derivative operation with respect to the same field and vice versa; it is so evident that the operator ν is self-adjoint. Hence a Hilbert space structure is embedded in $\mathscr{S}$; the positivity of the norm is assured by the requirement of the analyticity property of the functions $F^0[A^a{}_\mu(x)]$, $F^1{}_a{}^{\alpha(n)}[A^d{}_\mu(x)]$, $F^2{}_a{}^{\alpha(n)}{}_b{}^{\beta(m)}[A^d{}_\mu(x)],\ldots$, in Eq. (2.11). Furthermore, we get

$$[\boldsymbol{\delta}]^+(1)=\sum_{n=0}\Bigg\{\Bigg[[f^{abc}D_{\{\alpha(n)}A^a{}_{\mu\}}]\frac{\partial}{\partial D_{\{\alpha(n)}A^b{}_{\mu\}}(x)}+f^{abc}\mathscr{D}^{a*}{}_{\alpha(n)}G^*{}_{\mu\nu}(x)\frac{\partial}{\partial\mathscr{D}^{b*}{}_{\alpha(n)}G^*{}_{\mu\nu}(x)}\Bigg]\times\frac{\partial}{\partial C^c(x)}-\frac{1}{2}f^{abc}D_{\alpha(n)}C^a(x)\frac{\partial}{\partial D_{\alpha(n)}[C^b(x)C^c(x)]}\Bigg\},\tag{2.21}$$

$$[\boldsymbol{\delta}]^+(k)=\Bigg\{D_{\{\alpha(k-2)}A^a{}_{\mu\}}(x)+\sum_{n=0}\frac{(n+s-1)!}{n!(s-1)}\times f^{cba}D_{\{\alpha(n)}\,\partial_\mu A^c{}_{\nu\}}(x)\frac{\partial}{\partial D_{\{\alpha(n)}A^b{}_{\nu\}}(x)}\Bigg\}\times\frac{\partial}{\partial D_{\alpha(k-2)}\,\partial_\mu C^a(x)}$$
$$\equiv[\mathscr{N}^a{}_{\alpha(k-2),\mu}]^+(x)\frac{\partial}{\partial D_{\alpha(k-2)}\,\partial_\mu C^a(x)},\quad k=2,3,\ldots.\tag{2.22}$$

Our task is now to solve the system Eqs. (2.21) and (2.22); this will be done in the next section.

III. CALCULATIONS AND RESULTS

The first step is to solve the system

$$\boldsymbol{\delta}(1)\Delta(x)=0,\tag{3.1a}$$

$$[\boldsymbol{\delta}]^+(1)\Delta(x)=0,\tag{3.1b}$$

which implies

$$\{[\boldsymbol{\delta}]^+(1),\boldsymbol{\delta}(1)\}_+\Delta(x)=0.\tag{3.2}$$

The above condition can be better analyzed if we decompose the operator $\boldsymbol{\delta}(1)$ as

$$\boldsymbol{\delta}(1)=C^c(x)\boldsymbol{\delta}_c(1)+C^c(x)\Delta_c(1)+C^c(x)d_c(1)+\underline{\boldsymbol{\delta}(1)},\tag{3.3}$$

where

$$\boldsymbol{\delta}_c(1)=\sum_{n=0}\Bigg\{[f^{abc}D_{\{\alpha(n)}A^b{}_{\mu\}}(x)]\frac{\partial}{\partial D_{\{\alpha(n)}A^a{}_{\mu\}}(x)}+f^{abc}\mathscr{D}^{b*}{}_{\alpha(n)}G^*{}_{\mu\nu}(x)\times\frac{\partial}{\partial\mathscr{D}^{a*}{}_{\alpha(n)}G^*{}_{\mu\nu}(x)}\Bigg\},\tag{3.4a}$$

$$\Delta_c(1)=-\sum_{n=1}f^{abc}D_{\alpha(n)}C^b(x)\frac{\partial}{\partial D_{\alpha(n)}[C^a(x)]},\tag{3.4b}$$

$$d_c(1)=-\frac{1}{2}f^{abc}C^b(x)\frac{\partial}{\partial[C^c(x)]},\tag{3.4c}$$

$$\boldsymbol{\delta}(1)=-\frac{1}{2}\sum_{n>1;n>r>1}\frac{n!}{r!(n-r)!}\times[f^{abc}D_{\alpha(r)}C^b(x)D_{\alpha(n-r)}C^c(x)]\times\frac{\partial}{\partial D_{\alpha(n)}C^a(x)}.\tag{3.4d}$$

Tedious and lengthy calculations give

$$\Bigg\{[\boldsymbol{\delta}_c(1)+\Delta_c(1)+\mathbf{d}_c(1)]^+[\boldsymbol{\delta}_c(1)+\Delta_c(1)+\mathbf{d}_c(1)]\times\Delta(x)+[\mathbf{d}_c(1)]^+[\mathbf{d}_c(1)]\Delta(x)$$
$$-\frac{1}{2}\sum_{n,m>1;\,n>r>1;\,m>s>1}\Bigg\{\frac{n!}{r!(n-r)!}\frac{m!}{s!(m-s)!}\delta^m{}_n\delta^a{}_g\times f^{abc}[D_{\alpha(r)}C^b(x)D_{\alpha(n-r)}C^c(x)]\times f^{gde}\frac{\partial}{\partial D_{\alpha(m-s)}C^e(x)}\frac{\partial\Delta(x)}{\partial D_{\alpha(s)}C^d(x)}\Bigg\}$$
$$+\sum_{n>2}\frac{(2n!)}{(n!)^2}C_2D_{\alpha(n)}C^a(x)\frac{\partial\Delta(x)}{\partial D_{\alpha(n)}[C^a(x)]}\Bigg\}=0,\tag{3.5}$$

where

$$f^{mbc}f^{nbc}=\delta^{mn}C_2\tag{3.6}$$

and we have used

$$\begin{aligned}
&(C^c(x)\boldsymbol{\delta}_c(1),[C^a(x)\boldsymbol{\delta}_a(1)]^+)\\
&\quad=\boldsymbol{\delta}_c(1)^+\boldsymbol{\delta}_c(1)-[\boldsymbol{\delta}_c(1)^+\mathbf{d}_c(1)+\mathbf{d}_c(1)^+\boldsymbol{\delta}_c(1)],\\
&(C^c(x)\Delta_c(1),[C^a(x)\Delta_a(1)]^+)\\
&\quad=\Delta_c(1)^+\Delta_c(1)-[\Delta_c(1)^+\mathbf{d}_c(1)+\mathbf{d}_c(1)^+\Delta_c(1)],\\
&(C^c(x)\mathbf{d}_c(1),[C^a(x)\mathbf{d}_a(1)]^+)=2\,\mathbf{d}_c(1)^+\mathbf{d}_c(1),\\
&(C^c(x)\boldsymbol{\delta}_c(1),[C^a(x)\Delta_a(1)]^+)=2\Delta_a(1)^+\boldsymbol{\delta}_a(1),\\
&(C^c(x)\boldsymbol{\delta}_c(1),[C^a(x)\mathbf{d}_a(1)]^+)=2\,\mathbf{d}_a(1)^+\boldsymbol{\delta}_a(1),\\
&(C^c(x)\Delta_c(1),[C^a(x)\mathbf{d}_a(1)]^+)=2\,\mathbf{d}_a(1)^+\Delta_a(1),\\
&(C^c(x)\boldsymbol{\delta}_c(1),[\boldsymbol{\delta}(1)]^+)=(C^c(x)\Delta_c(1),[\boldsymbol{\delta}(1)]^+)\\
&\quad=(C^c(x)\mathbf{d}_c(1),[\boldsymbol{\delta}(1)]^+)=0.
\end{aligned}\tag{3.7}$$

The Hilbert space structure we have embedded gives the conditions

$$[\boldsymbol{\delta}_c(1)+\Delta_c(1)+\mathbf{d}_c(1)]\Delta(x)=0,\tag{3.8a}$$

$$[\mathbf{d}_c(1)]\Delta(x)=0,\tag{3.8b}$$

$$\sum_{m>1;m>s}\sum_{s>1}\frac{m!}{s!(m-s)!}f^{abc}\times\frac{\partial}{\partial D_{\alpha(m-s)}C^c(x)}\frac{\partial\Delta(x)}{\partial D_{\alpha(s)}C^b(x)}=0,\tag{3.8c}$$

$$\frac{\partial\Delta(x)}{\partial D_{\alpha(m)}C^a(x)}=0,\quad m=2,3,4,\ldots.\tag{3.8d}$$

Equations (3.8a) and (3.8b) say that $\Delta(x)$ can be written

$$\Delta(x) = F(D_{\{\alpha(n)}A^b{}_{\mu\}}, \mathscr{D}^{a*}{}_{\alpha(n)}G^*{}_{\mu\nu}(x), D_{\alpha(m)}C^a(x))T(C^a(x)), \tag{3.9}$$

where $T(C^a(x))$ and $F(D_{\{\alpha(n)}A^b{}_{\mu\}}, \mathscr{D}^{a*}{}_{\alpha(n)}G^*{}_{\mu\nu}(x), D_{\alpha(m)}C^a(x))$ are global invariant functions. The first depends only on the underivated $C^a(x)$ field, the second can contain a ghost field only if it is derivated. Furthermore, Eq. (3.8d) says that $\Delta(x)$ cannot depend on the space-time derivatives of the ghost field $C^a(x)$ of order greater than 1.

It is now a matter of tricks to show that $\Delta(x)$ does not depend on the first-order derivative either. Indeed the conditions (3.8c) and (3.8d) imply

$$\frac{\partial\Delta_a(1)\Delta(x)}{\partial[\partial_\mu C^a(x)]} = \Delta_a(1)\frac{\partial\Delta(x)}{\partial[\partial_\mu C^a(x)]} = 0, \tag{3.10}$$

since

$$\Delta_c(1)\Delta(x) = -f^{abc}\partial_\mu C^b(x)\frac{\partial\Delta(x)}{\partial[\partial_\mu C^a(x)]}. \tag{3.11}$$

On the other hand it is evident, from Eqs. (3.8a)–(3.10), that

$$\left[\frac{\partial\Delta_a(1),\Delta_b(1)}{\partial[\partial_\mu C^a(x)]}\right]\Delta(x) = 2f^{bar}\frac{\partial\Delta_r(1)\Delta(x)}{\partial[\partial_\mu C^a(x)]} = 0 = 2f^{bar}\Delta_r(1)\frac{\partial\Delta(x)}{\partial[\partial_\mu C^a(x)]} = 0 \tag{3.12}$$

and so

$$= \left[f^{rba}\frac{\partial}{\partial[\partial_\mu C^a(x)]},\Delta_b(1)\right]\Delta(x), \tag{3.13}$$

which, after a few calculations, gives

$$\frac{\partial\Delta(x)}{\partial[\partial_\mu C^a(x)]} = 0, \tag{3.14}$$

which is our desired result.

To sum up we have shown that the solution of the system Eqs. (3.1a) and (3.1b)

$$\Delta(x) = F(D_{\{\alpha(n)}A^b{}_{\mu\}}\mathscr{D}^{a*}{}_{\alpha(n)}G^*{}_{\mu\nu}(x))T(C^a(x)). \tag{3.15}$$

The next step is to solve the equations

$$\boldsymbol{\delta}(k)\Delta(x) = D_{\alpha(k-2)}\partial_\mu C^a(x)\mathscr{N}^a{}_{\alpha(k-2),\mu}(x)\Delta(x) = 0, \tag{3.16a}$$

$$[\boldsymbol{\delta}]^+(k)\Delta(x) = [\mathscr{N}^a{}_{\alpha(k-2),\mu}]^+(x)\frac{\partial\Delta(x)}{\partial D_{\alpha(k-2)}\partial_\mu C^a(x)} = 0 \quad (k = 2,3,...), \tag{3.16b}$$

that is,

$$([\boldsymbol{\delta}]^+(k),[\boldsymbol{\delta}](k))_+\Delta(x) = [\mathscr{N}^a{}_{\alpha(k-2),\mu}]^+(x) \times\mathscr{N}^a{}_{\alpha(k-2),\mu}(x)\Delta(x) = 0 \quad (k = 2,3,...), \tag{3.17}$$

since $\Delta(x)$ does not depend on the derivative of the ΦΠ field.

Now, the same tricks as before suggest the system of equations

$$\mathscr{N}^a{}_{\alpha(k-2),\mu}(x)\Delta(x) = 0 \quad (k = 2,3...). \tag{3.18}$$

The locality hypothesis suggests that $\Delta(x)$ can depend explicitly on the derivatives of the gauge field $D_{\{\alpha(n)}A^a{}_{\mu\}}(x)$ whose order has an upper bound (let us call it r); if it is so, the previous system reduces to $(r+1)$ equations,

$$\frac{\partial\Delta(x)}{\partial D_{\{\alpha(r)}A^a{}_{\mu\}}(x)} = 0, \tag{3.19a}$$

$$\frac{\partial\Delta(x)}{\partial D_{\{\alpha(r-1)}A^a{}_{\mu\}}(x)} + f^{abc}A^a{}_\nu(x)\frac{\partial\Delta(x)}{\partial D_{\{\alpha(r-1)}\partial_\mu A^a{}_{\nu\}}(x)} = 0, \tag{3.19b}$$

$$\frac{\partial\Delta(x)}{\partial D_{\{\alpha(r-2)}A^a{}_{\mu\}}(x)} + f^{abc}A^a{}_\nu(x)\frac{\partial\Delta(x)}{\partial D_{\{\alpha(r-2)}\partial_\mu A^a{}_{\nu\}}(x)} + rf^{abc}(\partial_0 A^a{}_\nu(x))\frac{\partial\Delta(x)}{\partial D_{\{\alpha(r-2)}\partial_\mu\partial_\sigma A^a{}_{\nu\}}(x)} = 0, \tag{3.19c}$$

$$\vdots$$

$$\frac{\partial\Delta(x)}{\partial D_{\{\alpha(r-k)}A^a{}_{\mu\}}(x)} + \sum_{n=0}^{k-1}\frac{(n+r-k+1)!}{n!(r-k+1)!}f^{abc}D_{\{\alpha(n)}A^b{}_{\nu\}}(x)\frac{\partial\Delta(x)}{\partial D_{\{\alpha(n+r-k)}\partial_\mu A^c{}_{\nu\}}(x)} = 0, \tag{3.19x}$$

$$\vdots$$

$$\frac{\partial\Delta(x)}{\partial_{\{\alpha(r-k)}\partial_\mu\partial_\sigma A^a{}_{\beta\}}(x)} + \frac{1}{3!}\sum_{n=0}^{r-3}\frac{(n+3)!}{n!}f^{abc}D_{\{\alpha(n)\}}A^b{}_{\nu\}}(x)\frac{\partial\Delta(x)}{\partial D_{\{\alpha(n)}\partial_\mu\partial_\sigma\partial_\beta A^c{}_{\nu\}}(x)} = 0, \tag{3.19u}$$

$$\frac{\partial\Delta(x)}{\partial_{\{\alpha}\partial_\mu A^a{}_{\sigma\}}(x)} + \frac{1}{2!}\sum_{n=0}^{r-2}\frac{(n+2)!}{n!}f^{abc}D_{\{\alpha(n)}A^b{}_{\nu\}}(x)\frac{\partial\Delta(x)}{\partial D_{\{\alpha(n)}\partial_\mu\partial_\sigma A^c{}_{\nu\}}(x)} = 0, \tag{3.19v}$$

$$\frac{\partial\Delta(x)}{\partial A^a{}_\sigma(x)} - + \sum_{n=0}^{r-1}(n+1)f^{abc}D_{\{\alpha(n)}A^b{}_{\nu\}}(x)\frac{\partial\Delta(x)}{\partial D_{\{\alpha(n)}\partial_\sigma A^c{}_{\nu\}}(x)} = 0. \tag{3.19w}$$

Now, Eq. (3.19a) says that $\Delta(x)$ does not depend on the r-order symmetrized derivative of the $A^a{}_\nu(x)$ field; if we substitute this result in the next equation it follows that it cannot depend on the $(r-1)$-order derivative either; if we continue, the trick goes on, and the outcome is that $\Delta(x)$ *cannot depend on any symmetrized derivative $D_{\{\alpha(n)}A^c{}_{\nu\}}(x)\,(0<n<r)$ of any order whatsoever.*

Therefore our final result is that

$$\Delta(x) = F(\mathscr{D}^{a*}{}_{\alpha(n)} G^*{}_{\mu\nu}(x)) T(C^a(x)). \tag{3.20}$$

We have to remark that, from a physical point of view, it should be more important to find the cohomology space of δ in the local functional space; however in this case, since the integration process has to be performed, *all the global properties* of space-time (dimensions, connectedness, etc.) have to be specified.

Anyhow our result is the first step to solve this problem, as pointed out in Refs. 1 and 6. Presumably, this result plus the result of Ref. 6 gives the proof for δ mod d, knowing that d is trivial (in general the "de Wilde" result).

ACKNOWLEDGMENT

We are grateful to Dr. P. Truini for a careful reading of the manuscript.

APPENDIX: SPECTRAL SEQUENCES: A BRIEF INTRODUCTION

In this appendix we give a brief introduction to the spectral sequence technique. For a more exhaustive treatment we refer to Ref. 4.

Let K be a differential complex with differential operator δ; i.e., K is an Abelian group and δ: $K \Rightarrow K$ is a group homomorphism such that $\delta^2 = 0$.

Suppose that in K there is a grading $K = \Sigma_{k\in Z} C^k$ and δ: $C^k \Rightarrow C^{k+1}$ increases the grading by 1.

The infinite sequence of subcomplexes

$$K = K(0) \supset K(1) \supset K(2) \supset \cdots \supset K(p)$$
$$\supset K(p+1) \supset \cdots \supset K(\infty) = 0 \tag{A1}$$

is called a filtration on K: this makes k a filtered complex. In our case the grading number is given by the eigenvalues of the ΦΠ ghosts and their space-time derivative counting operator ν, so the space $K(p)$ contains elements whose eigenvalue is greater than p.

Suppose now that the filtration has the property

$$\delta K(p) \subset K(p) \quad \text{for all } p. \tag{A2}$$

The operator δ will be graded with respect to ν as

$$[\nu,\delta] = \sum_{p>1} p\delta(p). \tag{A3}$$

Now define

$$\delta^{-1}K(p) = [x\in\delta^{-1}K(p) \text{ if } \delta x \in K(p)],$$
$$K^p{}_r = K(p) \cap \delta^{-1} K(p+r) \tag{A4}$$

and

$$Z^b{}_r = K(p) \quad \text{for } r<0, \tag{A5}$$

$$\delta Z^{p-r}{}_r = K(p) \cap \delta K(p-r). \tag{A6}$$

Obviously $Z^p{}_r$ contains $Z^{p+1}{}_{r-1}$ and $\delta Z^{p+1-r}{}_{r-1}$, so we can define

$$E^p{}_r = Z^p{}_r / [Z^{p+1}{}_{r-1} + \delta Z^{p+1-r}{}_{r-1}], \tag{A7}$$

$$E_r = \sum_p E^p{}_r. \tag{A8}$$

It is evident that if $x\in K(p)$ and $\delta x = 0$ then $x \in E^p{}_r$ for all r, and if $x\in K(p)$ and $x = \delta y$, then x is not an element of $E^p{}_r$ for r large enough.

Furthermore, δ maps $Z^p{}_r$ into $Z^{p+r}{}_r$ and $[Z^{p+r}{}_{r-1} + \delta Z^{p+1-r}{}_{r-1}]$ into $\delta Z^{p+r}{}_{r-1}$ and since $E^{p+r}{}_r = Z^{p+r}{}_r / [Z^{p+1+r}{}_{r-1} + \delta Z^{p+1}{}_{r-1}]$, δ will induce a differential d_r: $E^p{}_r \Rightarrow E^{p+r}{}_r$ whose cohomology space can be computed as follows: (1) the space $Z^p(E_r)$ of cycles in $E^p{}_r$ is defined by $x\in Z^p{}_r$ such that $\delta x \in [Z^{p+1+r}{}_{r-1} + \delta Z^{p+1}{}_{r-1}]$, i.e., $\delta x = \delta y + z$, with $y \in Z^{p+1}{}_{r-1}$, $z \in Z^{p+1+r}{}_{r-1}$. For $x = y + u$, we get $\delta u = z$, i.e., $u \in \delta^{-1} Z^{p+1+r}{}_{r-1}$, or, better $u \in k(p) \cap \delta^{-1} K(p+r+1) = Z^Z{}_{r+1}$, and, since $Z^{p+1}{}_{r-1}$ is in the "denominator" or $E^p{}_r$, we obtain

$$Z^p(E_r) = [Z^p{}_{r+1} + Z^{p+1}{}_{r-1}] / [Z^{p+1}{}_{r-1}$$
$$+ \delta Z^{p+1-r}{}_{r-1}] = E^p{}_r \cap Z^p{}_{r+1}; \tag{A9a}$$

(2) the space of coboundaries $B^p(E_r)$ in $E^p{}_r$ contains the elements $z \in \delta Z^{p-r}{}_r$, hence

$$B^p(E_r) = E^p{}_r \cap \delta Z^{p-r}{}_r$$
$$= [\delta Z^{p-r}{}_r + Z^{p+1}{}_{r-1}] / [Z^{p+1}{}_{r-1}$$
$$+ \delta Z^{p+1-r}{}_{r-1}] \tag{A9b}$$

and the cohomology space $H^p(E_r) = Z^p(E_r)/B^p(E_r)$ will have the form

$$H^p(E_r)$$
$$= [E^p{}_r \cap Z^p{}_{r+1}] / [E^p{}_r \cap \delta Z^{p-r}{}_r]$$
$$= [Z^p{}_{r+1} + Z^{p+1}{}_{r-1}] / [\delta Z^{p-r}{}_r + Z^{p+1}{}_{r-1}]$$
$$= Z^p{}_{r+1} / [Z^p{}_{r+1} \cap [\delta Z^{p-r}{}_r + Z^{p+1}{}_{r-1}]]$$
$$= Z^p{}_{r+1} / [\delta Z^{p-r}{}_r + Z^{p+1}{}_{r-1}] = E^p{}_{r+1} \subset E^p{}_r, \tag{A10}$$

since $DZ^{p-r}{}_r \subset Z^p{}_{r+1}$ and $Z^p{}_{r+1} \cap Z^{p+1}{}_{r-1} = Z^{p+1}{}_{r-1}$.

We have

$$\lim_{r\Rightarrow\infty} Z^p{}_r = K(p) \cap \delta^{-1} K(\infty) = K(p) \cap \delta^{-1} 0,$$

which represents the space of cocycles of $K(p)$,

$$\lim_{r\Rightarrow\infty} \delta Z^{p-r}{}_r = K(p) \cap \delta K(-\infty) = K(p) \cap \delta K,$$

which represents the space of coboundary of $K(p)$ and

$$E_\infty = \lim_{r\Rightarrow\infty} E_r = \lim_{r\Rightarrow\infty} \sum_p E^p{}_r$$

$$= \sum_p H_p(K)/H_{p+1}(K)$$

(which represents the graded cohomology group). Therefore the cohomology is reached first by the nested succession (in the index r) of the spaces $E^p{}_r$ and then summing in p in the sense of set theory.

Suppose now that K admits another filtration

$$K = K'(\infty) \supset \cdots \supset K'(p) \supset K'(p-1) \supset \cdots \supset K'(0) \tag{A11}$$

and $K'(p) = 0$ for $p < 0$ and that there exists a scalar product such that $K'(p)$ will be orthogonal to $K(q)$ for $p \neq q$.

In our case this is fulfilled if we use the same filtration as before if $K'(p)$ contains elements whose eigenvalue is less than or equal to p, and if we take the scalar product defined in Sec. III we can now prove the following theorem.

Theorem: If $x_p \in E^p{}_r$, then the same $x_p \in E^p{}_{r+1}$ if

$$\delta(r) x_p = 0, \tag{A12}$$

$$\delta^+(r) x_p = 0, \tag{A13}$$

where $\delta^+(r)$ is the adjoint of $\delta(r)$ derived by the adjointness operation induced by our scalar product.

Proof: If $x_p \in E^p{}_r$, from Eq. (A10) we shall get that the same $x_p \in E^p{}_{r+1}$ if (1) $x_p \in Z^p{}_{r+1}$ that is $x_p \in \delta^{-1} K(p+r+1)$, and this will imply $\langle \delta x_p | K'(p+r) \rangle = 0$, where $\langle 1 \rangle$ means scalar product [if we decompose δx_p and $K'(p+r)$ into their components belonging to different eigenspaces of the operator ν, then Eq. (A12) immediately follows]; and (2) x_p has to be orthogonal to all elements of the space $\delta Z^{p-r}{}_r$, that is, $\langle \delta^+ x_p | K(p-r) \rangle = 0$.

Equation (A13) is easily derived after the decomposition of $\delta^+ x_p$ and $K(p-r)$ into their ν eigenspaces.

Recalling now the definition of the space E_r we can use the above theorem to prove the following lemma.

Lemma: If $x \in E_r$, then the same $x \in E_{r+1}$ if

$$\delta(r) x = 0, \tag{A14}$$

$$\delta^+(r) x = 0, \tag{A15}$$

and the space E_∞ can be "formally" derived solving the above infinite system.

[1]C. Becchi, A. Rouet, and R. Stora, Ann. Phys. (NY) **98**, 287 (1976); R. Stora, in *New Developments in Quantum Field Theory*, edited by M. Le'vy and P. Mitter (Plenum, New York, 1977), p. 201.

[2]C. N. Yang and R. Mills, Phys. Rev. **96**, 191 (1954).

[3]J. Dixon, in Cohomology and Renormalization of Gauge Theories I, II, III (unpublished).

[4]J. Serre, Ann. Math **54**, 425 (1951); H. Godement, "Les fascieaux fibrée" in *Théorie des Faisceaux* (Hermann, Paris, 1952); see also R. Bott and L. Tu, *Differential Forms in Algebraic Topology* (Springer, New York, 1982).

[5]G. Bandelloni, Nuovo Cimento A **88**, 35 (1985); J. Math. Phys. **27**, 1128, 2251 (1986); **28**, 1875 (1987).

[6]M. Dobois-Violette, M. Talon, and C.-M. Viallet, Commun. Math. Phys. **102**, 105 (1986).

The Leibbrandt–Mandelstam prescription for general axial gauge one-loop integrals

P. Gaigg, M. Kreuzer, and M. Schweda
Institut für Theoretische Physik, Technische Universität Wien, Karlsplatz 13, A-1040 Wien/Vienna, Austria

O. Piguet
CERN, TH-Division, CH-1211 Genève, Switzerland

(Received 3 February 1987; accepted for publication 24 June 1987)

It is well known that, in doing light-cone gauge calculations, it is mandatory to regularize the unphysical $(q\,n)^{-\beta}$ poles by use of the Leibbrandt–Mandelstam prescription. This technique is also applied to general axial gauges and it is proved that it is a suitable regularization procedure for these gauges as well. In order to find the relation between the Leibbrandt prescription and the more familiar principal value prescription with its simpler Lorentz structure the temporal gauge limit $\mathbf{n}\to 0$ is performed (within dimensional regularization). Although this limit is found to be singular for multiple poles, the analytically regularized one-loop integrals agree with the results obtained within the principal value technique for the temporal gauge.

I. INTRODUCTION

Since Mandelstam proved the UV finiteness of $N=4$ Super Yang–Mills theories by means of the light-cone gauge,[1] this (very singular) gauge has become increasingly popular. Like the axial gauges the light-cone gauge is characterized by an arbitrary but constant vector n_μ. For the axial gauges n_μ need only satisfy $n^2\neq 0$, whereas for the light-cone gauge $n^2=0$. As a consequence of such gauges additional factors $(qn)^{-1}$ appear in the momentum-space propagator of the gauge field, and loop integrals become more intricate than in covariant gauges. A major problem is the consistent treatment of the unphysical singularity $(qn)^{-\beta}$. However, for axial gauges the principal value (PV) prescription has proved to be a well-suited (but not unique) way to implement power counting and unitarity.[2] It amounts to setting[3]

$$\frac{1}{(qn)^\beta}=\lim_{\varepsilon\to 0^+}\frac{1}{2}\left(\frac{1}{(qn+i\varepsilon)^\beta}+\frac{1}{(qn-i\varepsilon)^\beta}\right). \tag{1.1}$$

But for the light-cone gauge the PV prescription is afflicted with serious peculiarities, namely[4,5]: (a) some of the divergences created by one-loop corrections manifest themselves as double poles $(\omega-2)^{-2}$ (space-time dimension 2ω); and (b) the PV prescription gives rise to poles situated in the second and third quadrant of the complex q^0 plane which effectively prohibits Wick rotation and hence the application of standard power counting.

Because of these defects of the PV technique it had to be abandoned for the light-cone gauge. Instead of it Mandelstam and Leibbrandt independently introduced the so-called light-cone (LC) prescription[1,5]

$$\frac{1}{(qn)^\beta}=\lim_{\varepsilon\to 0^+}\left(\frac{(qn^*)}{(qn)(qn^*)+i\varepsilon}\right)^\beta,\quad \varepsilon>0, \tag{1.2}$$

where $n_\mu=(n_0,\mathbf{n})$ and $n^*_\mu=(n_0,-\mathbf{n})$, and proved that this LC prescription exhibits all the necessary items of a viable regularization of the $(qn)^{-\beta}$ poles. The vital point with the LC prescription is that two space-time directions n_μ and n^*_μ are singled out to regularize the $(qn)^{-\beta}$ pole *à la* Eq. (1.2), yielding well-behaved integrals at the price of a richer tensor structure of the integrals (terms proportional to n^*p, n^*n, ..., occur) and the appearance of nonlocalities in the divergent parts.

On the other hand, we find it desirable to investigate whether the LC prescription is applicable to axial gauges as well and, in doing so, to put the regularization of axial gauge poles and light-cone poles on equal footing. For arbitrary axial gauges (and therefore arbitrary n_μ and n^*_μ) this should be a straightforward procedure. However, in the temporal gauge $\mathbf{n}=0$ $(n_\mu=n^*_\mu)$ we have to expect difficulties as can be understood from

$$\lim_{\varepsilon\to 0^+}\frac{qn^*}{(qn)(qn^*)+i\varepsilon}=\mathrm{PV}\left(\frac{1}{qn}\right)-i\pi\,\mathrm{sgn}(qn^*)\delta(qn)\,, \tag{1.3}$$

which is obviously meaningless for the temporal gauge. Indeed, the limit $\mathbf{n}\to 0$ of Eq. (1.2) is singular for $\beta>1$ and some additional regularization is necessary. By analytic continuation of the exponent of the axial pole we obtain well-defined momentum integrals, which are identical to the PV results, as we will prove.

In the following only integrals with a factor $(qn^*)^\beta((qn)(qn^*)+i\varepsilon)^{-\beta}$ are analyzed in Minkowski space with Feynman parameters and dimensional regularization. More complicated expressions can be reduced by repeated use of the identity

$$\begin{aligned}&\frac{qn^*}{(qn)(qn^*)+i\varepsilon}\,\frac{(q+p)n^*}{(q+p)n(q+p)n^*+i\varepsilon}\\&=\frac{1}{pn+i\varepsilon pn^*}\left(\frac{qn^*}{(qn)(qn^*)+i\varepsilon}\right.\\&\qquad\left.-\frac{(q+p)n^*}{(q+p)n^*(q+p)n+i\varepsilon}\right).\end{aligned} \tag{1.4}$$

Remarkably enough one does not pick up additional contributions from δ functions, as it is the case for the corresponding formula holding in the PV technique:

 0022-2488/87/112781-05$02.50

$$\frac{1}{qn(q+p)n} = \frac{1}{pn}\left(\frac{1}{qn} - \frac{1}{(q+p)n}\right) + \pi^2\delta(nq)\delta(np) . \quad (1.5)$$

The paper is organized as follows: in Sec. II we derive the general formulas for LC-regularized one-loop integrals. Section III contains some new results on the light-cone gauge, whereas Sec. IV is devoted to the trickier business of the temporal gauge.

II. AXIAL ONE-LOOP INTEGRALS AND THE LC PRESCRIPTION

Due to Eq. (1.4) integrals to be computed in one-loop calculations can be reduced to [$g_{\mu\nu} = \text{diag}(1,-1,-1,-1)$, space-time dimension 2ω],

$$I(\alpha,\beta) := \int d^{2\omega}q\,(q^2+2pq-L+i\varepsilon)^{-\alpha}(qn^*)^\beta \times((qn^*)(qn)+i\eta)^{-\beta} , \quad \alpha\geqslant 1,\ \beta\geqslant 1,\ \varepsilon>0,\ \eta>0. \quad (2.1)$$

For further covenience we define

$$\bar{I}(\alpha,\beta;f(q)) := \int d^{2\omega}q\,(q^2+2pq\text{-}L+i\varepsilon)^{-\alpha} \times((qn^*)(qn)+i\eta)^{-\beta}f(q) . \quad (2.2)$$

Hence

$$I(\alpha,\beta) = [\Gamma(\alpha-\beta)/\Gamma(\alpha)](-D)^\beta\bar{I}(\alpha-\beta,\beta;1) , \quad (2.3)$$

where $D\equiv\frac{1}{2}n^*(\partial/\partial p)$. Note that regarding Wick rotation the LC prescription in a sense is much more natural than the PV prescription, because the denominator is positive semidefinite. Therefore in case of absolute convergence the above integral is well defined by analytic continuation to the Euclidean region if $L+p^2\geqslant 0$. To evaluate $\bar{I}(\alpha-\beta,\beta;1)$ we employ the conventional Feynman trick and the Euclidean identity

$$\int d^{2\omega}q\,(aq^2+2b(nq)+g(nq)^2+f)^{-\alpha} = \left(\frac{\pi}{a}\right)^\omega\sqrt{\frac{a}{C}}\frac{\Gamma(\alpha-\omega)}{F^{\alpha-\omega}} , \quad (2.4)$$

where

$$C = a+gn^2, \quad F = f-b^2n^2/C. \quad (2.4')$$

We obtain for $\bar{I}(\alpha-\beta,\beta;1)$,

$$\bar{I}(\alpha-\beta,\beta;1) = i\pi^\omega(-1)^\alpha\frac{\Gamma(\alpha-\omega)}{\Gamma(\alpha-\beta)\Gamma(\beta)}\int_0^1 dx \times(A\bar{A})^{-1/2}x^{-\beta}(1-x)^{\beta-1}\mathscr{L}^{\omega-\alpha} , \quad (2.5)$$

with the choice $n_\mu = (n_0,0,0,n_3)$ and the definitions

$$A = x+n_0^2(1-x) , \quad \bar{A} = x+n_3^2(1-x) , \quad \mathscr{L} = L+p^2+(1-x)\left(\frac{(p_3n_3)^2}{\bar{A}} - \frac{(p_0n_0)^2}{A}\right). \quad (2.6)$$

Now we apply the differential operator $(-D)^\beta$ to the integral (2.5), utilizing the general chain rule[6]

$$I(\alpha,\beta) \equiv \bar{I}(\alpha,\beta;(qn^*)^\beta) = i\pi^\omega\sum_{j=0}^{[\beta/2]}(-1)^{\alpha+j} \times\frac{\Gamma(\alpha-\omega+\beta-j)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\beta!}{(\beta-2j)!j!2^j} \times\int_0^1 dx\,x^{-\beta}(1-x)^{\beta-1} \times(A\bar{A})^{-1/2}\mathscr{L}^{\omega-\alpha-\beta+j}(D\mathscr{L})^{\beta-2j}(D^2\mathscr{L})^j , \quad (2.7)$$

where

$$D\mathscr{L} = x\left(\frac{n_0p_0}{A} + \frac{n_3p_3}{\bar{A}}\right), \quad D^2\mathscr{L} = \frac{1}{2}\frac{x^2n^2}{A\bar{A}}. \quad (2.7')$$

For arbitrary n_μ Eq. (2.7) leads to complicated generalized hypergeometric functions.[6] However, the divergent parts proportional to $(\omega-2)^{-1}$ can be integrated elementarily and are polynomials in p^2, pn, and pn^*. But—like in the light-cone gauge—the complete graphs may contain nonpolynomial parts due to the decomposition Eq. (1.4). Note that naive power counting is fulfilled and that $I(\alpha,\beta)$ is a regular function of n_μ for $n_0\neq 0$ and $n_3\neq 0$. Fortunately, for the most interesting limiting cases, namely the light-cone gauge ($n_3=n_0$) and the temporal gauge ($n_3=0$), $I(\alpha,\beta)$ can be evaluated in terms of hypergeometric functions of one variable: Due to its homogeneity of degree $-\beta$ in n we can simplify the integral (2.7) by setting $n_0=1$. The results for general n are recovered by substituting $n_3\to n_3/n_0$ and multiplication with $1/n_0^\beta$.

Because the rest of the paper will be dealing with these two gauges we now provide the appropriate values of Eq. (2.6):

$$\begin{aligned}
&n_0^2=1, \quad p_t^2 = p_1^2+p_2^2;\\
&A := x+n_0^2(1-x) = 1;\\
&\bar{A} := x+\mathbf{n}^2(1-x)\to\begin{cases}1, & \text{light-cone gauge,}\\ x, & \text{temporal gauge;}\end{cases}\\
&\mathscr{L} = L-p_t^2+xp_0^2-\frac{x}{\bar{A}}p_3^2\\
&\to\begin{cases}L-p_t^2+x(n^*p)(np), & \text{light-cone gauge,}\\ L-\mathbf{p}^2+xp_0^2 & \text{temporal gauge.}\end{cases}
\end{aligned} \quad (2.8)$$

III. THE LIGHT-CONE GAUGE

For the light-cone gauge the integral $I(\alpha,\beta)$ [Eq. (2.5)] can easily be calculated in terms of hypergeometric functions $F(a,b,c;z)$ (Ref. 6):

$$I(\alpha,\beta)=i(-1)^{\alpha}\pi^{\omega}\,\frac{\Gamma(\alpha+\beta-\omega)}{\Gamma(\beta+1)\Gamma(\alpha)}\times\left(\frac{2}{n^*n}\right)^{\beta}\frac{(n^*p)^{\beta}}{(L-p_t^2)^{\alpha+\beta-\omega}}\times F\left(1,\alpha+\beta-\omega,\beta+1;\ \frac{2}{n^*n}\,\frac{(n^*p)(np)}{p_t^2-L}\right). \quad (3.1)$$

Due to $D^2L:=\frac{1}{2}(x^2n^2/A\bar{A})=0$ on the light cone this integral is rendered more convergent than naive power counting would demand. Equation (3.1) is valid for arbitrary two-point functions in spontaneously broken gauge theories or QCD. Considering massless theories, i.e., $L+p^2=0$, yields the result

$$I(\alpha,\beta)=i(-1)^{\alpha}\pi^{\omega}\,\frac{\Gamma(\alpha+\beta-\omega)}{\Gamma(\alpha)\Gamma(\beta)(\omega-\alpha)}\times\frac{(n^*p)^{\beta}(n^*n/2)^{\alpha-\omega}}{[-(n^*p)(np)]^{\alpha+\beta-\omega}}. \quad (3.2)$$

This result is in agreement with special cases of this formula which have already been derived in the literature, e.g., see Ref. 5.

IV. THE TEMPORAL GAUGE

The central point of this paper is the investigation of the temporal gauge limit $n_\mu\to(1,\mathbf{0})$ within the LC prescription for the $q_0^{-\beta}$ poles. As already mentioned in the Introduction, switching over to the temporal gauge one encounters serious difficulties, due to the fact that the $q_0^{-\beta}$ pole is not completely regularized by the LC prescription. This feature is made explicit in the singular behavior of the momentum integrals at $n_3=0$, seen in the x integration at $x=0$. For the complete regularization of the $q_0^{-\beta}$ poles we will use analytic regularization.

As a first step we assume the exponents to be continuous; the asymptotic behavior of the momentum integrals for $n_3\to 0$ is then contained in the parameter integral

$$\int_0^1 dx\,x^{-\beta'}(x+n_3^2)^{-\alpha}=B(1-\beta',\alpha+\beta'-1)(n_3^2)^{1-\alpha-\beta}+\frac{1}{1-\alpha-\beta'}F(\alpha,\alpha+\beta'-1,\alpha+\beta';-n_3^2) \quad (4.1)$$

(note that the poles for $\alpha+\beta'=1$ cancel!). In order to find out for which β the integral Eq. (2.5) becomes singular we have to study $\bar{I}(\alpha,\beta';f(q))$ [Eq. (2.2)] for $f(q)=(qn^*)^{\beta}$. Evaluating this integral for $\beta=0,1$, and 2 we obtain

$$\bar{I}(\alpha,\beta';1)\sim\int_0^1 dx\,x^{-\beta'}(x+n_3^2)^{-1/2}=B\left(1-\beta',\beta'-\frac{1}{2}\right)n_3^{1-2\beta'}+\frac{2}{1-2\beta'}F\left(\frac{1}{2},\beta'-\frac{1}{2},\beta'+\frac{1}{2};-n_3^2\right), \quad (4.2)$$

$$\begin{aligned}\bar{I}(\alpha,\beta',qn^*)&\sim n_3\int_0^1 dx\,x^{1-\beta'}(x+n_3^2)^{-3/2}+\int_0^1 dx\,x^{1-\beta'}(x+n_3^2)^{-1/2}\\&=B\left(2-\beta',\beta'-\frac{1}{2}\right)(n_3^2)^{1-\beta'}+\frac{2n_3}{1\text{-}2\beta'}F\left(\frac{3}{2},\beta'-\frac{1}{2},\beta'+\frac{1}{2};-n_3^2\right)\\&\quad+B\left(2-\beta',\beta'-\frac{3}{2}\right)n_3^{3-2\beta'}+\frac{2}{3-2\beta'}F\left(\frac{1}{2},\beta'-\frac{3}{2},\beta'-\frac{1}{2};-n_3^2\right),\end{aligned} \quad (4.3)$$

$$\begin{aligned}\bar{I}(\alpha,\beta',(qn^*)^2)&\sim n_3^2\int_0^1 dx\,x^{2-\beta'}(x+n_3^2)^{-5/2}+\int_0^1 dx\,x^{2-\beta'}(x+n_3^2)^{-3/2}\\&=B\left(3-\beta',\beta'-\frac{1}{2}\right)n_3^{3-2\beta'}+\frac{2n_3^2}{1-2\beta'}F\left(\frac{5}{2},\beta'-\frac{1}{2},\beta'+\frac{1}{2};-n_3^2\right)\\&\quad+B\left(3-\beta',\beta'-\frac{3}{2}\right)n_3^{3-2\beta'}+\frac{2}{3-2\beta'}F\left(\frac{3}{2},\beta'-\frac{3}{2},\beta'-\frac{1}{2};-n_3^2\right).\end{aligned} \quad (4.4)$$

Thus $\bar{I}(\alpha,\beta';1)$, $\bar{I}(\alpha,\beta;qn^*)$, and $\bar{I}(\alpha,\beta';(qn^*)^2)$ are regular for $\beta'<\frac{1}{2}$, $\beta'\leqslant 1$, and $\beta'<\frac{3}{2}$, respectively, and

$$I(\alpha,\beta)\sim n_3^{1-\beta}, \quad (4.5)$$

so that $\lim_{n_3\to 0} I(\alpha,\beta)$ exists only for $\beta=1$.

For $\beta\geqslant 2$, $I(\alpha,\beta)$ can be defined by analytic continuation of $\bar{I}(\alpha,\beta',(qn^*)^{\beta})$ in β' at $\beta'=\beta$. As Hadamard's principal value is characterized by consistency with differentiation, which is guaranteed by analytic continuation, this amounts to taking the PV of $q_0^{\beta}(q_0^2+i\varepsilon)^{-\beta}$, which, in turn, is equivalent to the PV of $q_0^{-\beta}$ (this is easily "tested" with the basis $\{q_0^n|n\in\mathbb{N}\}$ of $L_2[-1,1]$). Thus we have proved that the LC prescription is not a complete regularization of multiple axial poles in the temporal gauge limit; further regularization by means of analytic continuation eventually is equivalent to the PV prescription.

In order to confirm this general argument we now turn to the evaluation of $I(\alpha,\beta)$ in the temporal gauge. We first define

$$c: = \alpha - \omega + \beta - j,$$
$$z: = \frac{p_0^2}{p^2 + L}, \quad \frac{z}{1-z} = \frac{p_0^2}{L - \mathbf{p}^2} \tag{4.6}$$

which we insert into Eq. (2.7),

$$I(\alpha,\beta) = \frac{i\pi^\omega(-1)^\alpha}{\Gamma(\alpha)\Gamma(\beta)} \sum_{j=0}^{[\beta/2]} \frac{\beta!(-1)^j}{(\beta-2j)!j!4^j} \times \int_0^1 dx\, x^{-1/2-j}(1-x)^{\beta-1} \times \frac{\Gamma(c)p_0^{\beta-2j}}{(L-\mathbf{p}^2)^c(1-xz/(1-z))^c}. \tag{4.7}$$

Using the integral representation of the hypergeometric function $F(a,b,c;z)$ (Ref. 7),

$$F(a,b,c;z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 dt\, t^{b-1} \times (1-t)^{c-b-1}(1-tz)^{-a} \tag{4.8}$$

and formula (9.132) of Ref. 6 we obtain

$$I(\alpha,\beta) = \frac{i\pi^\omega(-1)^\alpha}{\Gamma(\alpha)} \sum_{j=0}^{[\beta/2]} \frac{\beta!}{(\beta-2j)!} \times \frac{\pi}{j!4^j} \frac{\Gamma(c)}{\Gamma(\frac{1}{2}+j)\Gamma(\beta-j+\frac{1}{2})} \times \frac{p_0^{\beta-2j}}{(L+p^2)^c} F\left(\beta,c,\beta-j+\frac{1}{2};z\right). \tag{4.9}$$

The sum is calculated in the Appendix, yielding the final result

$$I(\alpha,\beta) = \frac{i\pi^\omega(-1)^\alpha}{n_0^\beta\Gamma(\alpha)} \begin{cases} \dfrac{\Gamma(\alpha+\beta/2-\omega)\Gamma(\frac{1}{2})}{\Gamma((\beta+1)/2)} \dfrac{F(\beta/2,\alpha+\beta/2-\omega,\frac{1}{2};z)}{(L+p^2)^{\alpha+\beta/2-\omega}}, & \beta \text{ even}, \\ \dfrac{\Gamma(\alpha+(\beta+1)/2-\omega)\Gamma(\frac{1}{2})}{\Gamma(\beta/2)} 2p_0 \dfrac{F((\beta+1)/2,\alpha+(\beta+1)/2-\omega,\frac{3}{2};z)}{(L+p^2)^{\alpha+(\beta+1)/2-\omega}}, & \beta \text{ odd}, \end{cases} \tag{4.10}$$

$n_\mu = (n_0,\mathbf{0})$, which is in complete agreement with the PV result of Konetschny.[3]

V. SUMMARY

In this paper we have proved that the LC prescription is a well-defined regularization of the *axial* gauge poles as well and derived the general formula for axial one-loop integrals within the LC prescription. We discussed the limiting cases of the light-cone gauge, where we found some new formulas, and the temporal gauge. For the latter the LC prescription does not regularize the $q_0^{-\beta}$ poles sufficiently: the limit $\mathbf{n}\to 0$ is singular. However, the analytically regularized LC results turn out to be identical to the results obtained within the PV technique. Hence, in a way, we put the regularization of the axial gauges and of the light-cone gauge on the same basis.

ACKNOWLEDGMENTS

We profited from a correspondence with G. Leibbrandt and a useful discussion with F. Vogl.

One of the authors (OP) was partially supported by the Swiss National Science Foundation.

APPENDIX: PROOF OF EQ. (4.10)

In order to prove the relation under scrutiny [Eq. (4.10)] we proceed from the identity

$$zF(a,b+1,c+1;z) = [c(c-1)/b(c-a)][F(a,b,c;z) - F(a-1,b,c-1;z)]. \tag{A1}$$

Accordingly we define a new function

$$\bar{F}(a,b,c;z) = [\Gamma(b)z^{c-a}/\Gamma(c)\Gamma(a-c+1)] \times F(a,b,c;z). \tag{A2}$$

Then the functions $\mathscr{F}(\beta,j)$ and the coefficients $c(\beta,j)$,

$$\mathscr{F}(\beta,j) = \bar{F}(\beta,a-j,\beta-j+\tfrac{1}{2};z),$$

$$c(\beta,j) = \begin{cases} \dfrac{\Gamma(\beta+1)}{\Gamma(\beta-2j+1)\Gamma(j+1)4^j}, & 0\leqslant j\leqslant \beta/2, \\ 0, & j\in\mathbb{Z}\setminus[0,\beta/2], \end{cases} \tag{A3}$$

fulfill the recursions

$$\mathscr{F}(\beta,j) = \mathscr{F}(\beta-1,j+1) - (\beta-j-\tfrac{3}{2})\mathscr{F}(\beta,j+1),$$
$$c(\beta,j) = c(\beta-2,j-1)(\beta-j-\tfrac{1}{2}) + c(\beta-2,j). \tag{A4}$$

Now we rewrite Eq. (4.9) in terms of $c(\beta,j)$ and $\mathscr{F}(\beta,j)$ and perform the sum using the recursion given above ($0\leqslant k\leqslant[\beta/2]$):

$$\sum_{j=0}^{[\beta/2]} c(\beta,j)\mathscr{F}(\beta,j) = \sum_{j=0}^{\infty} c(\beta-2k,j)\mathscr{F}(\beta-k,j+k)$$
$$= \sum_{j=0}^{\infty} c(\beta-2k-2,j-1)(\beta-j-\tfrac{1}{2}-2k)\mathscr{F}(\beta-k,j+k) + c(\beta-2k-2,j)$$
$$\times[\mathscr{F}(\beta-k-1,j+k+1) - (\beta-2k-j-\tfrac{3}{2})\mathscr{F}(\beta-k,j+k+1)]$$
$$= \sum_{j=0}^{\infty} c(\beta-2k-2,j)\mathscr{F}(\beta-k-1,j+k+1)$$
$$= c\left(\beta-2\left[\frac{\beta}{2}\right],0\right)\mathscr{F}\left(\beta-\left[\frac{\beta}{2}\right],\left[\frac{\beta}{2}\right]\right) = \mathscr{F}\left(\beta-\left[\frac{\beta}{2}\right],\left[\frac{\beta}{2}\right]\right). \quad \text{(A5)}$$

Inserting $\mathscr{F}(\beta,j)$ and $c(\beta,j)$ with $a=\alpha+\beta-\omega$ then yields Eq. (4.10)

[1]S. Mandelstam, Nucl. Phys. B **213**, 149 (1983).
[2]W. Kummer, Acta Phys. Austriaca **41**, 315 (1975); W. Konetschny and W. Kummer, Nucl. Phys. B **100**, 106 (1975); **108**, 397 (1976); J. Frenkel, Phys. Rev. D **13**, 2325 (1976).
[3]W. Konetschny, Phys. Rev. D **28**, 354 (1983); H. C. Lee and M. S. Milgram, Ann. Phys. (NY) **157**, 408 (1984).
[4]D. M. Capper, J. J. Dulwich, and M. J. Litvak, Nucl. Phys. B **241**, 463 (1984); H. C. Lee and M. S. Milgram, J. Math. Phys. **26**, 1793 (1985).
[5]G. Leibbrandt, Phys. Rev. D **29**, 1699 (1984); **30**, 2167 (1984); G. Leibbrandt and S. L. Nyeo, *ibid.* **33**, 3135 (1986).
[6]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1965).
[7]A. Erdelyi, W. Magnus, F. Oberhettinger, and F. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1.

Scalar potentials for vector fields in quantum electrodynamics

Eberhard E. Müller
Dublin Institute for Advanced Studies, School of Theoretical Physics, 10, Burlington Road, Dublin 4, Ireland

(Received 12 December 1985; accepted for publication 1 July 1987)

For an arbitrary vector field $\mathbf{F}$: $\mathbf{x}\in\mathbb{R}^3 \to \mathbf{F}(\mathbf{x})\in\mathbb{R}^3$, the representation $\mathbf{F} = \nabla\Phi + \mathbf{L}\Psi + \nabla\wedge\mathbf{L}\chi$, where Φ, Ψ, χ are scalar potentials, is used to disentangle the longitudinal and transversal degrees of freedom of the electromagnetic field. As a result the potentials for the electromagnetic field can be quantized restriction-free. Replacing the conventional vector potential in the Dirac equation by these Debye potentials provides a manifestly covariant and local description for the interaction of the quantized electromagnetic field with a quantized Dirac field.

I. INTRODUCTION

In the theory of electromagnetism the Helmholtz Theorem is commonly used to decompose a three-dimensional vector field $\mathbf{F}$: $\mathbf{x}\in\mathbb{R}^3 \to \mathbf{F}(\mathbf{x})\in\mathbb{R}^3$ as

$$\mathbf{F}(\mathbf{x}) = \nabla\Phi(\mathbf{x}) + \nabla\wedge\mathbf{A}(\mathbf{x}), \tag{1}$$

thereby introducing a scalar function Φ and a vector potential $\mathbf{A}$. However, there exist other decompositions in terms of three scalar potentials, the so-called Debye potentials.[1–3] Precisely stated one can decompose as follows.

Representation Theorem: Given a region $\Omega \subseteq \mathbb{R}^3\setminus\{0\}$, with regular boundary, and a C^3-vector field, $\mathbf{F}$: $\mathbf{x}\in\Omega \to \mathbf{F}(\mathbf{x})\in\mathbb{R}^3$. Let $\mathbf{L}$ denote the angular momentum operator. Then there exist three scalar functions Φ_F, Ψ_F, χ_F on Ω such that

$$\mathbf{F} = \nabla\Phi_F + \mathbf{L}\Psi_F + \nabla\wedge\mathbf{L}\chi_F. \tag{2}$$

Requiring Φ_F to vanish on the boundary, and ψ_F, χ_F not to contain spherically symmetric components, these functions are unique.

This theorem can be derived from the rigorous results in Ref. 1; a self-contained proof based on the Hodge decomposition for exterior differential forms[4] is given in the Appendix. Two proofs, one intuitive, the other formal, based on inverse operators, are given by Gray and Nickel.[2] The "gauge freedom" associated with decomposition (2) is the freedom to add spherically symmetric functions to the Debye potentials Ψ, χ. The requirement of spherical symmetry distinguishes this gauge from the usual vector potential gauge where the gradient of any scalar function can be added to the vector potential.

It is the purpose of this paper to show that the longitudinal and transversal degrees of freedom of the electromagnetic field disentangle if we use the scalar fields of (2) in the Maxwell equations. Without further assumptions, such as the Lorentz condition or Coulomb gauge, we arrive at wave equations for the transversal potentials. Therefore the transversal potentials can be quantized canonically without the notorious restrictions which are a main cause of the difficulties in quantum electrodynamics.[5] On the other hand, up to a time derivation, the gradient potentials of the electric field and the exterior current are identical. This shows that the longitudinal part of the Maxwell field belongs to the sources. With regard to the separation between transversal and longitudinal, it is interesting to notice that from a group theoretical point of view, where photons are defined by irreducible unitary rest mass zero representations of the Poincaré group, it is the introduction of the conventional vector potential that causes difficulties.[6]

The Debye potentials can be implemented in the Dirac equation in a Lorentz invariant form, thereby replacing the conventional vector potential. This provides a description for the interaction between photons and electrons that is both local and covariant.

II. MAXWELL EQUATIONS IN TERMS OF DEBYE POTENTIALS, CANONICAL QUANTIZATION

The Maxwell equation (in SI units) are given by

$$\nabla\wedge\mathbf{E} = -\dot{\mathbf{B}}, \tag{3a}$$

$$\nabla\wedge\mathbf{H} = \mathbf{J} + \dot{\mathbf{D}}, \tag{3b}$$

with the usual assumptions

$$\nabla\cdot\mathbf{D} = \rho, \tag{4a}$$

$$\nabla\cdot\mathbf{B} = 0, \tag{4b}$$

$$\epsilon\mathbf{E} = \mathbf{D}, \tag{5a}$$

$$\mathbf{B} = \mu\mathbf{H}, \tag{5b}$$

where the medium is supposed to be characterized by a dielectric constant ϵ and a constant magnetic permeability μ. Combining (5) and (3b) yields

$$\nabla\wedge\mathbf{B} = \mu\mathbf{J} + \epsilon\mu\dot{\mathbf{E}}. \tag{3b'}$$

Now we decompose the vector fields according to (2),

$$\mathbf{F} = \nabla\Phi_F + \mathbf{L}\Psi_F + \nabla\wedge\mathbf{L}\chi_F, \quad \mathbf{F} = \mathbf{E},\mathbf{B},\mathbf{J}, \tag{6}$$

under the tacit assumption that the gradient potentials vanish on the boundary of the region Ω under consideration which is supposed to exclude the origin, and to be sufficiently regular. A straightforward calculation implies the following form of the Maxwell equations:

$$\begin{aligned} &\dot{\Phi}_B = 0, \\ &\mathbf{L}^2\Psi_E = -\mathbf{L}^2\dot{\chi}_B, \\ &\nabla^2\mathbf{L}^2\chi_E = \mathbf{L}^2\dot{\Psi}_B, \end{aligned} \tag{7a}$$

$$\begin{aligned} &\dot{\Phi}_E = -(1/\epsilon)\Phi_J, \\ &\mathbf{L}^2\Psi_B = \mu\mathbf{L}^2\chi_J + \epsilon\mu\mathbf{L}^2\dot{\Psi}_E, \\ &\nabla^2\mathbf{L}^2\chi_B = -\mu\mathbf{L}^2\Psi_J - \epsilon\mu\mathbf{L}^2\dot{\Psi}_E. \end{aligned} \tag{7b}$$

The assumptions (4) are equivalent to

$$\nabla^2\Phi_E = (1/\epsilon)\,\rho, \tag{8a}$$

$$\Phi_B = 0. \tag{8b}$$

The remarkable result of this calculation is the separation of the gradient potentials Φ and the transversal potentials Ψ, χ. Without any further assumptions or restrictions, we arrive at a wave equation by eliminating Ψ_E and Ψ_B,

$$\Box \mathbf{L}^2\chi_E = \mu \mathbf{L}^2\dot{\chi}_J, \quad \Box := \nabla^2 - \epsilon\mu\,\partial_t^2, \tag{9a}$$

$$\Box \mathbf{L}^2\chi_B = -\mu \mathbf{L}^2\Psi_J. \tag{9b}$$

Equation (8a) and the first of Eqs. (7b) can be combined to give the continuity equation

$$\nabla^2\Phi_J + \dot{\rho} = 0, \tag{10}$$

which is nothing more than

$$\nabla\cdot\mathbf{J} + \dot{\rho} = 0. \tag{10'}$$

Equation (10) underlines that charge transport is exclusively related to the gradient potential of the current.

The angular momentum operator $\mathbf{L}$ annihilates spherically symmetric functions. Therefore, referring to the gauge principle, we can omit $\mathbf{L}^2$ in Eqs. (7) and (9). Thus (3a) is, up to a gauge, equivalent to

$$\dot{\Phi}_B = 0, \quad \Psi_E = -\dot{\chi}_B, \quad \nabla^2\chi_E = \dot{\Psi}_B, \tag{11a}$$

(3b′) is equivalent to

$$\begin{aligned} \dot{\Phi}_E &= -(1/\epsilon)\Phi_J, \\ \Psi_B &= \mu\chi_J + \epsilon\mu\dot{\chi}_E, \\ \nabla^2\chi_B &= -\mu\Psi_J - \epsilon\mu\dot{\Psi}_E, \end{aligned} \tag{11b}$$

and (9) is equivalent to

$$\Box\chi_E = \mu\dot{\chi}_J, \tag{12a}$$

$$\Box\chi_B = -\mu\Psi_J. \tag{12b}$$

The formulation (7) [resp. (12)] of the Maxwell equations exhibits a structure crucial for the quantization procedure. By the first of Eqs. (7b), the electric gradient field Φ_E coincides with a matter field. Hence its possible quantization is subject to a quantum theory of matter. Moreover we want the triviality of the magnetic gradient field Φ_B due to the first of Eqs. (7a) [resp. Eq. (8b)] to be maintained in the quantization. Therefore we have to omit Φ_E and Φ_B from the quantization of the free electromagnetic field. In the wave equation (9) [resp. (12)] the inhomogeneities refer only to matter while the homogeneous solutions precisely exhaust the free electromagnetic field. Therefore the quantization of the free electromagnetic field is achieved by the canonical quantization of the solutions of the homogeneous wave equations

$$\Box\chi_\alpha = 0, \quad \alpha = E', B', \quad \mathbf{E}' := \epsilon^{1/2}\mathbf{E}, \quad \mathbf{B}' := \mu^{-1/2}\mathbf{B}, \tag{13}$$

and presents no problem at all. General quantized solutions of (9) involve the quantization of the inhomogeneities as well which has to originate in quantum theory of matter.

We sketch the canonical quantization of (13). Here $(\nabla^2 - \epsilon\mu\,\partial_t^2)\chi_\alpha = 0$ can be derived from the Lagrange density

$$\mathscr{L}(\dot{\chi}_\alpha, \nabla\chi_\alpha) = \frac{1}{2}\sum_{\alpha = E', B'} (\epsilon\mu\dot{\chi}_\alpha^2 - (\nabla\chi_\alpha)\cdot(\nabla\chi_\alpha)) \tag{14}$$

by the Euler–Langrange equation

$$\partial_t \frac{\partial\mathscr{L}}{\partial\dot{\chi}_\alpha} + \nabla\cdot\frac{\partial\mathscr{L}}{\partial(\nabla\chi_\alpha)} = 0. \tag{15}$$

The canonical momentum with respect to χ_α is

$$\pi_\alpha = \frac{\partial\mathscr{L}}{\partial\dot{\chi}_\alpha} = \epsilon\mu\dot{\chi}_\alpha, \tag{16}$$

and the Hamiltonian density

$$\begin{aligned} \mathscr{H} &= \frac{1}{2}\sum_{\alpha = E', B'} \left(\frac{\pi_\alpha^2}{\epsilon\mu} + (\nabla\chi_\alpha)\cdot(\nabla\chi_\alpha)\right) \\ &= \frac{1}{2}\sum_{\alpha = E', B'} \left(\frac{\pi_\alpha^2}{\epsilon\mu} - \chi_\alpha\nabla^2\chi_\alpha\right). \end{aligned} \tag{17}$$

A Fourier transformation with respect to the time variable makes (13) assume the form of an eigenvalue equation,

$$-\nabla^2 u_k = \epsilon\mu\omega_k^2 u_k. \tag{18}$$

We specify a boundary value problem on Ω, Dirichlet data say, such that the Laplacian is self-adjoint. Now let us expand $\chi_\alpha(\mathbf{x},t)$ with respect to an orthonormal basis $\{u_k\}$ of eigenfunctions of $-\nabla^2$:

$$\begin{aligned} \chi_\alpha(\mathbf{x},t) = \sum_k (2\epsilon\mu\omega_k)^{1/2}\{ & b_{k\alpha}u_k(\mathbf{x})e^{-i\omega_k t} \\ & + b_{k\alpha}^* u_k^*(\mathbf{x})e^{-i\omega_k t}\}; \end{aligned} \tag{19}$$

the orthogonality may be given with respect to the inner product

$$\begin{aligned} \langle f,g\rangle &:= \int_\Omega f^*(\mathbf{x})g(\mathbf{x})d^3\mathbf{x}, \\ \langle u_k, u_{k'}\rangle &= \delta_{kk'}. \end{aligned} \tag{20a}$$

We assume a nontrivial time dependence in (19), $\omega_k > 0$ for all k (see final section). Now insert (19) in (16) and (17). Observe that both $u_{k'}$ and u_k^* are eigenfunctions to the same eigenvalue in (18) such that by

$$\begin{aligned} 0 &= \langle u_k, (-\nabla^2)u_{k'}^*\rangle - \langle -\nabla^2 u_k, u_{k'}^*\rangle \\ &= \epsilon\mu(\omega_{k'}^2 - \omega_k^2)\langle u_k, u_{k'}^*\rangle, \end{aligned}$$

we have

$$\langle u_k, u_{k'}^*\rangle = 0, \quad \omega_k \neq \omega_{k'}. \tag{20b}$$

Hence the Hamiltonian density

$$\mathscr{H}_p = \tfrac{1}{2}\omega_p(b_p^* b_p + b_p b_p^*), \quad p := (k,\alpha), \quad \alpha = E', B', \tag{21}$$

follows straightforwardly. Canonical quantization consists in postulating the commutation relations

$$[b_p, b_p^*]_- = \hbar\delta_{pp'}, \quad [b_p, b_{p'}]_- = 0. \tag{22}$$

Therefore the Hamiltonian density is

$$\mathscr{H}_p = \omega_p b_p^* b_p + \tfrac{1}{2}\hbar\omega_p. \tag{23}$$

We can rearrange the decomposition of the electromagnetic field into the poloidal potentials $\chi_{E'}, \chi_{B'}$, to represent the helicity

$$\chi_K := (1/\sqrt{2})(\chi_{B'} + i^\kappa\chi_{E'}), \quad \kappa = +1, -1. \tag{24a}$$

This transformation leaves the commutation relations (22) unaffected, i.e., they are valid for

$$q=(k,\kappa), \quad \kappa=+1,-1, \tag{24b}$$

$$b_q := (1/\sqrt{2})(b_{k,B}+i^\kappa b_{k,E}), \tag{24c}$$

$$[b_q, b^*_{q'}] = \hbar\delta_{qq'}, \quad [b_q, b_{q'}] = 0. \tag{25}$$

The Hamiltonian density with respect to helicity is

$$\mathscr{H}_q = \omega_q b^*_q b_q + \tfrac{1}{2}\hbar\omega_q. \tag{26}$$

III. DEBYE POTENTIALS AND DIRAC EQUATION

In this section we describe the interaction between the electromagnetic field and a Dirac field in terms of Debye potentials. This is achieved by rewriting the conventional vector potential in terms of Debye potentials. The resulting description is manifestly covariant.

Let us assume the poloidal current potential χ_J to vanish. Then the magnetic induction **B** can be written as the second of Eqs. (7b),

$$\mathbf{B} = \mathbf{L}\Psi_B + \nabla\wedge\mathbf{L}\chi_B$$
$$= \nabla\wedge(\mathbf{x}i\,\partial_t\,\epsilon_0\mu_0\chi_E + \mathbf{L}\chi_B). \tag{27}$$

(In this section we set $\epsilon=\epsilon_0$, $\mu=\mu_0$.) Therefore

$$\mathbf{x}i\,\partial_t\,\epsilon_0\mu_0\chi_E + \mathbf{L}\chi_B \tag{28}$$

is a vector potential. The expression $\mathbf{x}i\,\partial_t$ in the first term of (28) can be considered as part of a Lorentz boost generator

$$\mathbf{K} := \mathbf{x}(1/i)\sqrt{\epsilon_0\mu_0}\,\partial_t + (1/\sqrt{\epsilon_0\mu_0})t(1/i)\nabla. \tag{29}$$

The addition of the gradient of a scalar function does not affect the **B** field. We use this freedom to define a vector potential

$$\mathbf{C} := -\mathbf{K}\sqrt{\epsilon_0\mu_0}\chi_E + \mathbf{L}\chi_B. \tag{30}$$

With respect to parity the electric field **E** transforms odd while the magnetic induction **B** transforms even.[7] Hence χ_B is a pseudoscalar field, and χ_E a scalar field. Therefore **C** transforms as a proper three-vector. Its transformation behavior is dictated by the vector operators **K**, **L** which represent the Lie algebra of Lorentz generators. Under spatial rotations, **C** transforms form invariant,

$$\begin{pmatrix}\mathbf{K}\sqrt{\epsilon_0\mu_0}\chi_E(t,\mathbf{x})\\ \mathbf{L}\chi_B(t,\mathbf{x})\end{pmatrix} \to \begin{pmatrix}\mathbf{K}'\sqrt{\epsilon_0\mu_0}\chi'_E(t',\mathbf{x}')\\ \mathbf{L}'\chi'_B(t',\mathbf{x}')\end{pmatrix}, \tag{31a}$$

while under infinitesimal Lorentz boosts

$$\begin{pmatrix}\mathbf{K}\sqrt{\epsilon_0\mu_0}\chi'_E(t,\mathbf{x})\\ \mathbf{L}\chi_B(t,\mathbf{x})\end{pmatrix} \to \begin{pmatrix}i\mathbf{L}'\sqrt{\epsilon_0\mu_0}\chi'_E(t',\mathbf{x}')\\ i\mathbf{K}'\chi'_B(t',\mathbf{x}')\end{pmatrix}. \tag{31b}$$

This follows from the commutation relations for the infinitesimal generators of the Lorentz group,

$$[L_k, L_l] = i\epsilon_{klm}L_m, \tag{32a}$$

$$[L_k, K_l] = i\epsilon_{klm}K_l, \tag{32b}$$

$$[K_k, K_l] = -i\epsilon_{klm}L_m. \tag{32c}$$

The transition from the vector potential (28) to the vector potential (30) conflicts with the gauge freedom associated with decomposition (2); the boost generator **K** does not necessarily annihilate functions dependent on $|\mathbf{x}|$. The scalar functions that are invariant under Lorentz transformations, and that are annihilated by the infinitesimal generators **K**, **L**, are

$$f(|\mathbf{x}|)\in\mathbb{C}, \quad |\mathbf{x}| := (t^2/\epsilon_0\mu_0 - \mathbf{x}^2)^{1/2}. \tag{33}$$

Thus the gauge freedom associated with decomposition (30) is the freedom to add to the potentials χ_E, χ_B functions which are spherically symmetric with respect to the Minkowski space.

Now we use the three-vector potential **C** to form a four-vector potential

$$(A_\mu) := \begin{pmatrix}\varphi\\ \mathbf{C}\end{pmatrix}, \tag{34}$$

where the scalar-valued potential φ is defined by

$$\mathbf{E} =: -\nabla\varphi - \dot{\mathbf{C}} \tag{35}$$
$$= \nabla\Phi_E + \mathbf{L}\Psi_E + \nabla\wedge\mathbf{L}\chi_E$$
$$= \nabla\Phi_E + \nabla\wedge(\mathbf{K}(\chi_B/\sqrt{\epsilon_0\mu_0}) + \mathbf{L}\chi_E); \tag{36}$$

in the last line we employed the Bianchi identity [the second of Eqs. (11a)], and

$$\mathbf{L}\sqrt{\epsilon_0\mu_0}\,\partial_t = -\nabla\wedge\mathbf{K}. \tag{37}$$

By (35) we get (up to a constant)

$$\varphi = -\Phi_E - i(x^\mu\partial_\mu + 2)\chi_E. \tag{35'}$$

We would like to use the four-vector potential (34) in the Dirac equation. Because of our assumption of a vanishing poloidal current potential χ_J which was essential to get the four-vector potential in the form (34), we have to recur to the variational principle for the Lagrangian of the interaction of Dirac field and electromagnetic field to ensure the consistency of the use of A_μ in the form (34).

With the familiar notation

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{38}$$

the purely electromagnetic part of the Lagrangian is given as

$$\mathscr{L}'_{\text{em}} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}$$
$$= \tfrac{1}{2}\epsilon_0\mathbf{E}^2 - (1/2\mu_0)\mathbf{B}^2$$
$$= \tfrac{1}{2}\epsilon_0(-\Phi_E\nabla^2\Phi_E + \chi_E\Box\mathbf{L}^2\chi_E)$$
$$- (1/2\mu_0)\chi_B\Box\mathbf{L}^2\chi_B. \tag{39}$$

Clearly the variational principle for this Lagrangian reproduces the free case of Eqs. (8) and (9). But neither these equations nor the Lagrangian (39) are manifestly covariant. The presence of the operators $\mathbf{L}^2$ and ∇^2 in (39) [resp. (8) and (9)] reflects the noncovariance of the Debye decomposition (2). Under an infinitesimal Lorentz boost $\mathbf{L}^2$ transforms into $-\mathbf{K}^2$. Therefore the requirement of manifest covariance imposes the substitution of $\mathbf{L}^2\chi_E$ by $-\mathbf{K}^2\chi_E$, the conjugation $(\chi_E, \chi_B)\to(i\chi_B, i\chi_E)$ under infinitesimal boost, and similarly the substitution of ∇^2 by $\Box$. This yields the following Lagrangian:

$$\mathscr{L} = \bar{\psi}(\gamma^0(-(\hbar/i)\partial_t - e\Phi_E)$$
$$+ \gamma\cdot(-(\hbar c/i)\nabla - e(\mathbf{M}+\mathbf{M}^*) - mc^2))\psi$$
$$+ \tfrac{1}{2}\epsilon_0(-\Phi_E\Box\Phi_E + (\mathbf{M}+\mathbf{M}^*)\cdot\Box(\mathbf{M}+\mathbf{M}^*)), \tag{40}$$

$$\mathbf{M}: = (- \mathbf{L} + i\mathbf{K})\chi, \quad \chi: = \tfrac{1}{2}(i\chi_E - (1/\sqrt{\epsilon_0\mu_0})\chi_B).$$

Now variation with respect to $(\mathbf{M} + \mathbf{M}^*)$, and Φ_E implies the following set of coupled equations of motion:

$$(\gamma^0(- (\hbar/i)\,\partial_t - e\Phi_E) + \boldsymbol{\gamma}\cdot(- (\hbar c/i)\nabla - e(\mathbf{M} + \mathbf{M}^*)) - mc^2)\psi = 0, \quad (41)$$

$$\Box\Phi_E = - (e/\epsilon_0)\bar{\psi}\gamma^0\psi, \quad (42a)$$

$$\Box(\mathbf{M} + \mathbf{M}^*) = - (e/\epsilon_0)\bar{\psi}\boldsymbol{\gamma}\psi, \quad (42b)$$

The canonical quantization of the electromagnetic potentials χ_E, χ_B (resp. χ, χ^*) has to be adapted to the inhomogeneous wave equations, and presents no problem. Again we quantize restriction-free. The quantization of the Dirac field is carried out as usual.

IV. DISCUSSION

(1) The use of the electromagnetic Debye potentials in the Dirac equation provides a manifestly Lorentz-covariant description that at the same time preserves locality under the quantization. Moreover the Lagrange density (40) is invariant under gauge transformations of the potentials χ_E, χ_B, i.e., addition of scalar functions dependent on $((\epsilon_0\mu_0)^{-1}t^2 - x^2)^{1/2}$ only.

(2) The presence of the angular momentum operator in (9) is physically significant. It prevents the photon fields χ_E, χ_B from being spherically symmetric. This guarantees a nonvanishing helicity. Simultaneously this excludes the possibility of a nontrivial field for the eigenvalue zero in the Helmholtz equation (18) for the free case, which otherwise would emerge in the infinite volume limit, or for Neumann data, etc. Thereby the conceptual problem is circumvented that a rest mass zero particle with kinetic energy zero cannot exist.

The gauge freedom associated with decomposition (2) introduces the nonuniqueness of the ground state.

(3) It is obvious that the Debye–Hodge decomposition (2) and the decomposition (30) deserve a completely group-theoretical analysis. This will be done in a forthcoming paper.[8] Here we only remark that the operators $\mathbf{L}$, $i^{-1}(-\nabla^2)^{-1/2}\nabla\wedge\mathbf{L}$ satisfy the algebraic relations (32).[9]

Moreover it is clear that the Dirac equation with the scalar valued electromagnetic potentials will considerably simplify QED calculations. This is the subject of current work.[10]

ACKNOWLEDGMENTS

I am indebted to Professor J. T. Lewis, Professor L. O'Raifeartaigh, and Professor A. Wipf for conversations on the above topic. Preliminary work has been done at the ETH Zürich during the author's postdoctoral research with Professor H. Primas.

The author was supported in part by Deutsche Forschungsgemeinschaft.

APPENDIX: INGREDIENTS OF HODGE'S DECOMPOSITION THEOREM

We use Hodge's decomposition theorem for exterior differential forms of degree p on a closed orientable n-dimensional Riemannian manifold Ω, in the form given by Ref. 4. It states that every C^2 p-form α can be decomposed uniquely into a sum of three forms

$$\alpha = \alpha_1 + \alpha_2 + \alpha_3, \quad (A1)$$

α_1 being exact, α_2 coexact, and α_3 harmonic; i.e., there exists a $(p-1)$-form β_1 with

$$\alpha_1 = d\beta_1 \quad (A2)$$

and a $(p+1)$-form γ_2 with

$$\alpha_2 = \delta\gamma_2: = (d\gamma_2^*)^* \quad (A3)$$

while, for $\Delta: = (- 1)^{np}\delta d + (- 1)^{np+n}\,d\delta$,

$$\Delta\alpha_3 = 0. \quad (A4)$$

If the manifold Ω is two dimensional and α a one-form, the forms β_1 and γ_2^* are scalars. If in addition Ω is a sphere the first Betti number is zero, and α_3 vanishes. Hence we have the special situation that the decomposition of the two-component vector α is given in terms of two scalar fields β_1 and γ_2^*. This implies the following lemma (see Ref. 1).

Lemma 1: Let Ω be a two-dimensional sphere and $\mathbf{F}_t$ a C^2-vector field on Ω. Let ∇_t denote the gradient on Ω, and $\hat{r}$ the unit vector at $\mathbf{r}\in\Omega\subset\mathbb{R}^3$ perpendicular to Ω. Then there exist functions S and T on Ω such that

$$\mathbf{F}_t = \nabla_t S + \hat{r}\wedge\nabla_t T. \quad (A5)$$

Here $\nabla_t S$ gives the exact one-form $d\beta_1$, and $\hat{r}\wedge\nabla_t T$ the coexact one-form $\delta\gamma_2 = \delta(T d\Omega)$. (The $d\Omega$ denotes the obvious two-dimensional differential.)

To apply this lemma for a three-dimensional vector field $\mathbf{F}$: $\mathbf{x}\in\mathbb{R}^3\to\mathbf{F}(\mathbf{x})\in\mathbb{R}^3$, we use radial and tangential coordinates:

$$\mathbf{F}(\mathbf{x}) = : F_1(\mathbf{x})\hat{x} + \mathbf{F}_t(\mathbf{x}). \quad (A6)$$

So excluding the origin we relate $\mathbf{x}\in\mathbb{R}^3$ one-to-one to $(x,\hat{x})\in\mathbb{R}^+\times\Omega$, $x: = |\mathbf{x}|$, with Ω the unit sphere, and $\hat{x}: = \mathbf{x}/x$. The F_1 is a scalar function on the domain of $\mathbf{F}$, and $\mathbf{F}_t(\mathbf{x})$ a tangent vector to Ω at $\hat{x}$. In these coordinates the gradient is given as

$$\nabla_{\mathbf{x}} = : \frac{\partial}{\partial x}\hat{x} + (\nabla_{\mathbf{x}})_t. \quad (A7)$$

Proof of the Representation Theorem: We may always add a spherically symmetric function of Ψ_F, χ_F without changing (2). Let us assume Φ_F to vanish on the boundary, and Ψ_F, χ_F not to be spherically symmetric. We use the Poisson formula[11]

$$\mathbf{F}(\mathbf{x}) = \nabla^2\int_\Omega K(\mathbf{x},\mathbf{y})\mathbf{F}(\mathbf{y})d^3\mathbf{y}, \quad (A8)$$

where the kernel K is a fundamental solution of the Laplace equation, and abbreviate $\mathbf{G}(\mathbf{x}): = \int_\Omega K(x,y)\mathbf{F}(\mathbf{y})d^3\mathbf{y}$. Employing the decompositions (A6) and (A5), and some vector identities, we have

$$\begin{aligned}\mathbf{F}(\mathbf{x}) &= \nabla^2\mathbf{G}(\mathbf{x})\\ &= \nabla(\nabla\cdot\mathbf{G}(\mathbf{x})) - \nabla\wedge(\nabla\wedge\mathbf{G}(\mathbf{x}))\\ &= \nabla(\nabla\cdot\{G_1(\mathbf{x})\hat{x} + \nabla_t S(\mathbf{x}) + \hat{x}\wedge\nabla_t T(\mathbf{x})\})\\ &\quad - \nabla\wedge(\nabla\wedge\{G_1(\mathbf{x})\hat{x} + \nabla_t S(\mathbf{x}) + \hat{x}\wedge\nabla_t T(\mathbf{x})\})\\ &= \nabla(\nabla\cdot(G_1(\mathbf{x})\hat{x}) + \nabla_t^2 S(\mathbf{x}))\end{aligned}$$

$$- \nabla \wedge (\nabla \wedge \hat{x} G_1(\mathbf{x})) + \nabla \wedge \left(\nabla \wedge \hat{x} \frac{\partial}{\partial x} S(\mathbf{x})\right)$$
$$- \nabla \wedge (\nabla \wedge \{\hat{x} \wedge \nabla T(\mathbf{x})\})$$
$$= \nabla(\nabla \cdot (G_1(\mathbf{x})\hat{x}) + \nabla_t^2 S(\mathbf{x}))$$
$$+ \nabla \wedge \mathbf{L} \frac{i}{x} \left(G_1(\mathbf{x}) - \frac{\partial}{\partial x} S(\mathbf{x})\right)$$
$$+ \mathbf{L} \nabla^2 \left(i \frac{T(\mathbf{x})}{x}\right).$$

This proves the existence of the Debye potentials.

Suppose now

$$\mathbf{F} = \nabla \Phi + \mathbf{L} \Psi + \nabla \wedge \mathbf{L} \chi = \nabla \Phi' + \mathbf{L} \Psi' + \nabla \wedge \mathbf{L} \chi'.$$

Then straightforward vector calculations imply

$$\nabla^2 \Phi = \nabla \cdot \mathbf{F}, \quad \mathbf{L}^2 \Psi = \mathbf{L} \cdot \mathbf{F}, \quad \mathbf{L}^2 \nabla^2 \chi = -(\mathbf{L} \wedge \nabla) \cdot \mathbf{F},$$

hence

$$\nabla^2(\Phi - \Phi') = 0, \quad \mathbf{L}^2(\Psi - \Psi') = 0,$$
$$\mathbf{L}^2 \nabla^2 (\chi - \chi') = -x^2 \nabla_t^4 (\chi - \chi') = 0,$$

and therefore

$$\Phi = \Phi', \quad \Psi = \Psi', \quad \chi = \chi'. \qquad \text{Q.E.D.}$$

[1]C. H. Wilcox, "Debye potentials," J. Math. Mech. **6**, 167 (1957). (See additional references therein.)

[2]C. G. Gray, "Multipole expansions of electromagnetic fields using Debye potentials," Am. J. Phys. **46**, 169 (1978); C. G. Gray and B. G. Nickel, "Debye potential representation of vector fields," Am. J. Phys. **46**, 735 (1978). (See additional references therein.)

[3]P. Debye, "Der Lichtdruck auf Kugeln von beliebigem Material," Ann. Phys. (Leipzig) **30**, 57 (1909); W. W. Hansen, "A new type of expansion in radiation problems," Phys. Rev. **47**, 139 (1935); C. J. Bouwkamp and H. B. G. Casimir, "On multipole expansions in the theory of electromagnetic radiation," Physica **20**, 539 (1954); A. Nisbet, "Hertzian electromagnetic potentials and associated gauge transformations," Proc. R. Soc. London Ser. A **231**, 250 (1955); G. Backus, "A class of self sustaining dissipative spherical dynamos," Ann. Phys. (NY) **4**, 372 (1958); S. C. Snowdon, "Theory of multipole radiation," J. Math. Phys. **2**, 719 (1961); W. B. Campbell and T. Morgan, "Debye potentials for the gravitational field," Physica **53**, 264 (1971); E. G. P. Rowe, "Decomposition of vector fields by scalar potentials," J. Phys. A **12**, 145 (1979); D. K. Monroe; "Lorentz transformation of Debye potentials," J. Math. Phys. **25**, 1787 (1984).

[4]P. Bidal and G. de Rham, "Les formes différentielles harmonique," Commun. Math. Helv. **19**, 1 (1956).

[5]F. Strocchi, "Gauge problem in quantum field theory," Phys. Rev. **162**, 1429 (1967); Phys. Rev. D **2**, 2334 (1970).

[6]U. H. Niederer and L. O'Raifearthaigh, "Realizations of the unitary representations of the inhomogeneous space-time groups II," Fortschr. Phys. **22**, 131 (1974).

[7]J. D. Jackson, *Class Electrodynamics* (Wiley, New York, 1975).

[8]In collaboration with Professor G. Parravicini.

[9]A. C. T. Wu, "Debye scalar potentials for the electromagnetic fields," Phys. Rev. D **34**, 3109 (1986).

[10]E. E. Müller, "Asymptotic coupling of matter and radiation," work in progress.

[11]F. John, *Partial Differential Equations* (Springer, New York, 1982).

Electrohydrodynamic stability of two superposed elasticoviscous liquids in plane Couette flow

Nabile T. M. Eldabe
Department of Mathematics, Hail T. T. Junior College, P. O. Box 1818, Saudi Arabia

(Received 13 March 1986; accepted for publication 27 May 1987)

The electrohydrodynamic stability of two superposed elasticoviscous liquids (Oldroyd's liquid B) in plane Couette flow is studied. The system is stressed by a normal electric field. The analysis includes all possible modes of perturbations. The eigenvalue problem governing the stability of the flow against wave formation is formulated and solved for small values of wave numbers. It is found that the electric field cannot only destabilize the flow but it can also stabilize the flow for certain values of dielectric ratio, depth ratio, viscosity ratio, elasticity ratio, density ratio, Froude number, Reynolds number, and the electric potential of the plate.

I. INTRODUCTION

The stability of non-Newtonian fluid has received increasing interest due to its technological applications such as petroleum industries, packed bed reactors, and many other industries.

The investigation of the stability of non-Newtonian fluids has been undertaken by Yih,[1] Chun and Schwabs,[2] Gupta,[3] Lai,[4] and Fan Chun.[5,6] The stability of two superposed fluids has been studied by Yih,[7] Hickox,[8] Kao,[9] Li,[10] and Shivamoggi.[11]

Here we aim to study the effect of a normal electric field on the stability of two superposed elasticoviscous liquids in plane Couette flow. Li's[10] perturbation technique is used in the following analysis.

II. THE PRIMARY FLOW

The system considered here consists of two finite homogeneous dielectric non-Newtonian fluids u and l between two parallel walls having the upper boundary moving with a constant velocity U_0 and of V_0 potential. The lower boundary is stationary and of zero potential in order to produce a normal electric field (see Fig. 1.). Let $(u^*)^k$, $(v^*)^k$, $(w^*)^k$ denote the velocity components in the x^*, y^*, z^* directions, respectively, where x^*, y^*, and z^* are Cartesian coordinates. The superscript k is u for the upper liquid and l for lower liquid. The prototype of liquid designed by Oldroyd[12] is considered. For this liquid, the rheological equations are

$$S_{ik} = -\pi\delta_{ik} + p_{ik} \tag{1}$$

and

$$p^{ik} + \lambda_1 \frac{d_c p^{ik}}{dt} = 2\eta_0\left(h^{ik} + \lambda_2 \frac{d_c h^{ik}}{dt}\right), \tag{2}$$

in which S_{ik} is the stress tensor, $\pi = p - (\frac{1}{2})\epsilon^* E^2$, p is the hydrostatic pressure, ϵ^* is the dielectric constant, E is the electric field, δ_{ik} is the Kronecker delta, $h^{ik} = (u^*_{i,k} + u^*_{k,i})/2$ is the rate-of-strain tensor, η_0 is a coefficient of viscosity, λ_1 is the relaxation time, and λ_2 is the retardation time. The coefficients η_0, λ_1, and λ_2 $(<\lambda_1)$ are all positive. The symbol d_c/dt denotes the convective derivative of a tensor quantity in relation to the fluid in motion. For a contravariant tensor B^{ik}

$$\frac{d_c B^{ik}}{dt} = \frac{\partial B^{ik}}{\partial t} + B^{ij}_{,m} V^m - B^{mj} V^i_{,m} - B^{im} V^j_{,m}\,. \tag{3}$$

No volume charges are present in the bulk of the fluids. Also because of the continuity of the electric field, no surface charges are present at the interfaces in the equilibrium state and will therefore vanish during the perturbation.[13] Due to the potential difference there exists a normal electric field whose form in the unperturbed state can be determined from Poisson's equation

$$\nabla\cdot\epsilon^*\mathbf{E} = 0\,. \tag{4}$$

We assume that the quasistatic approximation[14] is valid for the problem and therefore the electric field is a curl-free vector,

$$\nabla\wedge\mathbf{E} = 0\,. \tag{5}$$

Therefore we can define an electric potential function Q such that

$$\mathbf{E} = -\nabla Q\,. \tag{6}$$

From Eq. (4)

$$\mathbf{E} = T\mathbf{e}_j\,. \tag{7}$$

Then potential Q is

$$Q = -Ty + D\,, \tag{8}$$

where T and D are constants, since the rheological equations and equations of motion governing the upper liquid and the

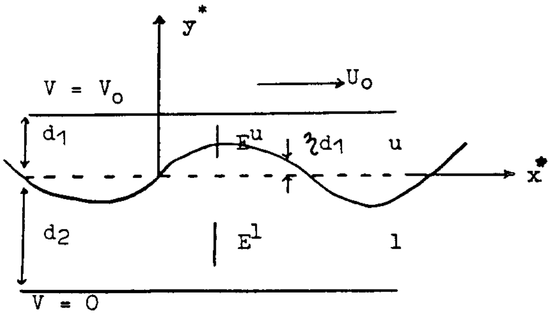

FIG. 1. Definition sketch.

lower liquid will be in the same form. At this stage we drop the superscript for convenience, and consider the steady undisturbed flow

$$u^* = \bar{u}(y^*), \quad v^* = 0, \quad w^* = 0.$$

For this flow, the constitutive equations can be written as follows:

$$\begin{aligned}
&\bar{p}_{11} + \lambda_1\left(\frac{\partial \bar{p}_{11}}{\partial t} + \frac{\partial \bar{p}_{11}}{\partial x^*}\bar{u} - 2\bar{p}_{12}\frac{\partial \bar{u}}{\partial y^*}\right) = -2\eta_0\lambda_2\left(\frac{\partial \bar{u}}{\partial y^*}\right)^2, \\
&\bar{p}_{12} + \lambda_1\left(\frac{\partial \bar{p}_{12}}{\partial x^*}\bar{u} - \bar{p}_{22}\frac{\partial \bar{u}}{\partial y^*}\right) = \eta_0\frac{\partial \bar{u}}{\partial y^*}, \\
&\bar{p}_{13} + \lambda_1\left(\frac{\partial \bar{p}_{13}}{\partial x^*}\bar{u} - \bar{p}_{23}\frac{\partial \bar{u}}{\partial y^*}\right) = 0, \\
&\bar{p}_{22} + \lambda_1\frac{\partial \bar{p}_{22}}{\partial x^*}\bar{u} = 0, \\
&\bar{p}_{23} + \lambda_1\frac{\partial \bar{p}_{23}}{\partial x^*}\bar{u} = 0, \\
&\bar{p}_{33} + \lambda_1\frac{\partial \bar{p}_{33}}{\partial x^*}\bar{u} = 0.
\end{aligned} \tag{9}$$

The equations of motion can be written as

$$\begin{aligned}
0 &= -\frac{\partial \bar{\pi}}{\partial x^*} + \frac{\partial \bar{p}_{11}}{\partial x^*} + \frac{\partial \bar{p}_{21}}{\partial y^*} + \frac{\partial \bar{p}_{31}}{\partial z^*}, \\
0 &= -\frac{\partial \bar{\pi}}{\partial y^*} + \frac{\partial \bar{p}_{12}}{\partial x^*} + \frac{\partial \bar{p}_{22}}{\partial y^*} + \frac{\partial \bar{p}_{32}}{\partial z^*} - \rho g, \\
0 &= -\frac{\partial \bar{\pi}}{\partial z^*} + \frac{\partial \bar{p}_{13}}{\partial x^*} + \frac{\partial \bar{p}_{23}}{\partial y^*} + \frac{\partial \bar{p}_{33}}{\partial z^*}.
\end{aligned} \tag{10}$$

Equations (9) and (10) admit the stress components of primary flow to be

$$\begin{aligned}
&\bar{p}_{13} = 0, \quad \bar{p}_{22} = 0, \quad \bar{p}_{23} = 0, \\
&\bar{\pi} = \bar{\pi}(y^*), \quad \bar{p}_{11} = \bar{p}_{11}(y^*), \quad \bar{p}_{12} = \bar{p}_{12}(y^*).
\end{aligned} \tag{11}$$

We make all quantities nondimensional by letting

$$\begin{aligned}
&x^* = xd_1, \quad y^* = yd_1, \quad \bar{u} = UU_0, \\
&\bar{\pi} = \rho_1 U_0^2 \pi^*, \quad \epsilon = \epsilon^* d_1, \\
&E = (\rho^u d_1)^{1/2} U_0 E^*, \quad \bar{p}_{ij} = \rho^u U_0^2 P_{ij}, \quad t = d_1\tau/U_0.
\end{aligned}$$

The nondimensional forms of the first two equations of (9) and (10) are then

$$P_{11}^i - 2M_1^i P_{12}^i \frac{dU^i}{dy} = -2\beta^i \frac{M_2^i}{R}\left(\frac{dU^i}{dy}\right)^2, \tag{12a}$$

and

$$P_{12}^i = \frac{\beta^i}{R}\frac{dU^i}{dy}, \tag{12b}$$

$$0 = \frac{dP_{21}^i}{dy}, \tag{12c}$$

$$0 = -\frac{d\pi^{*i}}{dy} - \gamma^i F^{-2}, \tag{12d}$$

in which $R = U_0\rho^u d_1/\eta_0^u$ is the Reynolds number, $M_1^i = U_0\lambda_1^i/d_1$, $M_2^i = U_0\lambda_2^i/d_1$, and $F^2 = U_0^2/gd_1$ is the Froude number. The superscript i is taken to be u and l for the upper and the lower layer of liquids, respectively,

$$\beta^u = 1, \quad \beta^l = m_\eta, \quad m_\eta = \eta_0^l/\eta_0^u \tag{13}$$

is the ratio of viscosity and

$$\gamma^u = 1, \quad \gamma^l = \nu, \quad \nu = \rho^l/\rho^u \tag{14}$$

is the ratio of density. From Eqs. (12b) and (12c) the equations governing the primary flow can be obtained. These are

$$\frac{d^2U^u}{dy^2} = 0 \quad \text{and} \quad \frac{d^2U^l}{dy^2} = 0. \tag{15}$$

They are subject to the boundary condition that U^u is equal to a specified U_0 on the upper boundary and U^l is zero on the lower boundary. Also, U^u and U^l and P_{12}^u and P_{12}^l must be continuous at the interface. Equations (15) can be solved to yield the solutions

$$U^u = a_1 y + b \quad \text{and} \quad U^l = a_2 y + b \tag{16}$$

in which

$$\begin{aligned}
&a_1 = m_\eta/(m_\eta + n), \quad b = n/(m_\eta + n), \\
&a_2 = 1/(m_\eta + n), \quad n = d_2/d_1.
\end{aligned}$$

Equations (12a)–(12d) become

$$\begin{aligned}
&P_{12}^u = a_1/R, \quad P_{11}^u = (2/R)(M_1^u - M_2^u)a_1^2, \\
&\frac{d\pi^{*u}}{dy} = -F^{-2}, \quad P_{12}^l = \frac{a_2 m_\eta}{R} = P_{12}^u, \\
&P_{11}^l = \frac{2m_\eta}{R}(M_1^l - M_2^l)a_2^2, \quad \frac{d\pi^{*l}}{dy} = -\nu F^{-2}.
\end{aligned} \tag{17}$$

III. PERTURBATION EQUATIONS

We assume that the interface $y = 0$ is slightly disturbed such that

$$y = \eta = \delta \exp[i\alpha(x - c\tau)], \tag{18}$$

where δ is a smallness parameter, α indicates the wave number, and c accounts for the complex phase velocity of the disturbance. The electric field, the potential, the velocity, the pressure, the stress tensor, and the rate-of-strain tensor are then

$$\begin{aligned}
&\mathbf{E} = \mathbf{E}_0 + \mathbf{E}', \quad Q = Q_0 + Q', \quad u^* = U + u', \\
&v^* = v', \quad \pi^* = \Pi + \pi', \quad p_{11} = P_{11} + \sigma_{11}, \\
&p_{12} = P_{12} + \sigma_{12}, \quad p_{22} = \sigma_{22}, \quad h_{11} = h'_{11}, \\
&h_{12} = H_{12} + h'_{12}, \quad h_{22} = h'_{22},
\end{aligned} \tag{19}$$

in which the strain components are nondimensionalized by the unit U_0/d_1, and σ_{11}, σ_{12}, σ_{22}, and the quantities denoted by a prime indicate the small perturbation from the equilibrium state. The linearized equations of motion are

$$\begin{aligned}
&\frac{\partial u'}{\partial \tau} + U\frac{\partial u'}{\partial x} + v'\frac{\partial U}{\partial y} = -\frac{\partial \Pi'}{\partial x} + \frac{\partial \sigma_{11}}{\partial x} + \frac{\partial \sigma_{21}}{\partial y}, \\
&\frac{\partial v'}{\partial \tau} + U\frac{\partial v'}{\partial x} = -\frac{\partial \Pi'}{\partial y} + \frac{\partial \sigma_{21}}{\partial x} + \frac{\partial \sigma_{22}}{\partial y}.
\end{aligned} \tag{20}$$

The linearized equations of state are

$$\begin{aligned}
\sigma_{11} + M_1\bigg[&\frac{\partial \sigma_{11}}{\partial \tau} + \frac{\partial \sigma_{11}}{\partial x}U + v'\frac{\partial P_{11}}{\partial y} \\
&- 2\left(\frac{\partial u'}{\partial x}P_{11} + \frac{\partial u'}{\partial y}P_{12} + \sigma_{12}\frac{dU}{dy}\right)\bigg]
\end{aligned}$$

$$= \frac{2}{R} h'_{11} + \frac{2M_2}{R} \left[\frac{\partial h'_{11}}{\partial \tau} + \frac{\partial h'_{11}}{\partial x} U - \left(\frac{\partial u'}{\partial y} + 2h'_{12} \right) \frac{dU}{dy} \right],$$

$$\sigma_{12} + M_1 \left(\frac{\partial \sigma_{12}}{\partial \tau} + \frac{\partial \sigma_{12}}{\partial x} U + v' \frac{\partial P_{12}}{\partial y} - \sigma_{22} \frac{dU}{dy} - \frac{\partial v'}{\partial x} P_{11} \right)$$
$$= \frac{2}{R} h'_{12} + \frac{2M_2}{R} \left(\frac{\partial h'_{12}}{\partial \tau} + \frac{\partial h'_{12}}{\partial x} U + \frac{v'}{2} \frac{d^2 U}{dy^2} - h'_{22} \frac{dU}{dy} \right), \tag{21}$$

$$\sigma_{22} + M_1 \left(\frac{\partial \sigma_{22}}{\partial \tau} + \frac{\partial \sigma_{22}}{\partial x} U - 2 \frac{\partial v'}{\partial x} P_{12} \right)$$
$$= \frac{2}{R} h'_{22} + \frac{2M_2}{R} \left(\frac{\partial h'_{22}}{\partial \tau} + \frac{\partial h'_{22}}{\partial x} U - \frac{\partial v'}{\partial x} \frac{dU}{dy} \right).$$

It is clear that the linearization of Eqs. (4) and (5) leads to the following:

$$\nabla^2 Q' = 0, \quad \text{where } \mathbf{E}' = -\nabla Q'. \tag{22}$$

In addition to these, we have the equation of continuity

$$\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} = 0. \tag{23}$$

In order to retain the meaning of R and nondimensionalize the pressure by the same unit $\rho^u U_0^2$ for both layers of liquid, a factor $1/\nu$ will arise on the right-hand sides of Eqs. (20) and another factor m_η will arise in Eqs. (21) for those terms that contain the Reynolds number R, when the lower layer is considered.

Equation (21) permits the use of a streamfunction in terms of which

$$u' = \frac{\partial \psi}{\partial y}, \quad v' = -\frac{\partial \psi}{\partial x}. \tag{24}$$

As a result of perturbation, Q', Π', ψ, σ_{11}, σ_{12}, and σ_{22}, in view of the dependence given by Eq. (18), may have the form

$$(Q', \Pi', \psi, \sigma_{11}, \sigma_{12}, \sigma_{22}) = [\hat{Q}(y), f(y), \phi(y), F_1(y), F_2(y)] \exp[i\alpha(x - c\tau)]. \tag{25}$$

Substituting from the first equation of Eqs. (25) into Eq. (22), the solution of the differential equation is

$$Q'^{u,l} = (A^{u,l} \exp(\alpha y) + B^{u,l} \exp(-\alpha y)) \times \exp[i\alpha(x - c\tau)], \tag{26}$$

where $A^{u,l}$ and $B^{u,l}$ are constants that are to be evaluated by making use of the appropriate boundary conditions. Substituting Eqs. (25) and (24) into Eqs. (20) we eliminate Π' and obtain, for the two layers in turn,

$$i\alpha[(U^u - c)(\phi^{u''} - \alpha^2 \phi^u)] = i\alpha F_1^{u\prime} + F_2^{u\prime\prime} + \alpha^2 F_2^u - i\alpha F_3^{u\prime},$$
$$i\alpha[(U^l - c)(\phi^{\prime\prime l} - \alpha^2 \phi^l)] = i\alpha F_1^{\prime l} + F_2^{\prime\prime l} + \alpha^2 F_2^l - i\alpha F_3^{\prime l}. \tag{27}$$

Similarly, substituting Eqs. (17) into Eqs. (21) for the two layers, in turn,

$$F_1^i [1 + i\alpha M_1^i (U^i - c)] = M_1^i \{ i\alpha P_{11}^{\prime i} \phi^i + 2[i\alpha P_{11}^i \phi^{\prime i} + P_{12}^i \phi^{\prime\prime i} + F_2^i U^{\prime i}] \} + 2i\alpha\beta^i \phi^i / R - (2\beta^i M_2^i / R) \times [\alpha^2 \phi^{\prime i}(U^i - c) + U^{\prime i}(2\phi^{\prime\prime i} + \alpha^2 \phi^i)],$$

$$F_2^i [1 + i\alpha M_1^i (U^i - c)] = M_1^i [i\alpha (P_{12})^{\prime i} \phi^i + F_3^i U^{\prime i} + \alpha^2 \phi^i P_{11}^i] + (\beta^i / R)(\phi^{\prime\prime i} + \alpha^2 \phi^i) + (i\alpha\beta^i M_2^i / R) \times [(\phi^{\prime\prime i} + \alpha^2 \phi^i)(U^i - c) - \phi^i U^{\prime\prime i} + 2\phi^{\prime i} U^{\prime i}], \tag{28}$$

$$F_3^i [1 + i\alpha M_1^i (U^i - c)] = 2M_1^i P_{12}^i \alpha^2 \phi^i - 2i\alpha\beta^i \phi^{\prime i}/R + (2\beta^i M_2^i \alpha^2 / R) \times [\phi^{\prime i}(U^i - c) - U^{\prime i} \phi^i],$$

in which the superscript i is taken to be u and l as before for the upper layer and the lower layer of fluids, respectively.

IV. BOUNDARY CONDITIONS

From Eqs. (7) and (8) the electric fields and the potentials in the equilibrium state are

$$\mathbf{E}_0^{u,l} = C^{u,l} \mathbf{e}_j \quad \text{and} \quad Q_0^{u,l} = -C^{u,l} y + D^{u,l}. \tag{29}$$

The fields and the potentials are subjected to the following conditions.

(a) The electric potential is continuous at the interface.

(b) Since there are no surface charges on the interface, the normal electric displacement is continuous across the interface.

Applying the above conditions and simplifying we get

$$E_0^l = -V_0 \epsilon^u / (\epsilon^u n + \epsilon^l),$$
$$E_0^u = -V_0 \epsilon^l / (\epsilon^u n + \epsilon^l),$$
$$Q_0^l = V_0 \epsilon^u (y + n) / (\epsilon^u n + \epsilon^l), \tag{30}$$

and

$$Q_0^u = V_0 \epsilon^l [y + n\epsilon^u / \epsilon^l] / (\epsilon^u n + \epsilon^l),$$

where the subscript 0 will refer to the equilibrium state.

(c) The continuity of the electric potential at the perturbed interfaces leads to

$$Q'^u = Q'^l \quad \text{at} \quad y = \delta \exp[i\alpha(x - c\tau)]$$

and

$$Q'^u = 0 \quad \text{at } y = 1, \quad Q'^l = 0 \quad \text{at } y = -n,$$

noting that

$$Q^{u,l}(\eta) = Q_0^{u,l}(0) + \eta \left. \frac{\partial Q_0^{u,l}}{\partial y} \right|_{y=0} + Q'^{u,l}(0).$$

(d) The continuity of the normal electric displacement across the perturbed interface is, namely,

$$\epsilon^u (\mathbf{E}^u \cdot \mathbf{N}) = \epsilon^l (\mathbf{E}^l \cdot \mathbf{N}) \quad \text{at } y = \eta,$$

where $\mathbf{N}$ is the unit vector normal to the interface, $\mathbf{N} = -\delta i\alpha \exp[i\alpha(x - c\tau)]\mathbf{e}_i + \mathbf{e}_j$.

Note that

$$E_0^{u,l}(\eta) = E_0^{u,l}(0) + \eta \left. \frac{\partial E_0^{u,l}}{\partial y} \right|_{y=0}.$$

Hence, from Eqs. (26) and conditions (c) and (d), we obtain

$$Q'^u = V_0\epsilon^l(\epsilon^u - \epsilon^l)\delta[\exp(\alpha y) - \exp(\alpha(2-y))] \times \exp[i\alpha(x - c\tau)]/(\epsilon^u n + \epsilon^l) \times [(\epsilon^l - \epsilon^u) - (\epsilon^l + \epsilon^u)\exp(2\alpha)] \qquad (31a)$$

and

$$Q'^l = V_0\epsilon^u(\epsilon^u - \epsilon^l)(1 + \exp(2\alpha)) \times\delta[\exp(\alpha y) - \exp(-\alpha(2n + y))] \times\exp[i\alpha(x - c\tau)]/(\epsilon^u n + \epsilon^l) \times[(\epsilon^l - \epsilon^u) - (\epsilon^l + \epsilon^u)\exp(2\alpha)] \times(1 - \exp(-2\alpha n)).$$

(e) The normal components of the stress tensor Π_{ij} should be discontinuous at the interface by the surface tension T^*, where

$$\Pi_{ij} = -(\Pi + \epsilon E_k E_k/2)\delta_{ij} + \epsilon E_i E_j + p_{ij},$$

i.e.,

$$(N_i\Pi_{iy})^u - (N_i\Pi_{iy})^l = -S\delta_{ij}\nabla^2\eta,$$

in which $S = T^*/\rho^u U_0^2 d_1$ and again, variables are evaluated at $y = 0$. On substitution from Eqs. (17) and (20) the normal stress condition (e) can be rewritten as

$$-\alpha(c'\phi'^u + a_1\phi^u) - \alpha F_1^u + iF_2'^u + \alpha F_3^u + \alpha\nu(c'\phi'^l + a_2\phi^l) + \alpha F_1^l - iF_2'^l - \alpha F_3^l = \alpha[(\nu - 1)F^{-2} + \alpha^2 S]\delta - \alpha^2 V_0^2\epsilon^u\epsilon^l(1 + \exp(2\alpha)) \times\delta(\epsilon^u - \epsilon^l)^2(1 + \exp(2\alpha))/(\epsilon^u n + \epsilon^l) \times[\epsilon^l - \epsilon^u - (\epsilon^l + \epsilon^u)\exp(2\alpha)]. \qquad (31b)$$

(f) The shear stress is continuous at the perturbed interface, this condition yields

$$F_2^u = F_2^l - i\alpha V_0^2\epsilon^u\epsilon^l(\epsilon^u - \epsilon^l)\delta/(\epsilon^u n + \epsilon^l) \quad \text{at } y = 0. \qquad (31c)$$

(g) On the boundaries, the zero normal velocity and nonslip condition demand that

$$\phi^u(1) = 0, \qquad (31d)$$

$$\phi^u(1) = 0, \qquad (31e)$$

$$\phi^l(-n) = 0, \qquad (31f)$$

$$\phi^l(-n) = 0. \qquad (31g)$$

(h) The continuity of v' at the interface demands that

$$\phi^u(0) = \phi^l(0). \qquad (31h)$$

(i) The kinematic boundary condition at the interface is

$$\left(\frac{\partial\eta}{\partial\tau} + U^u\frac{\partial\eta}{\partial x}\right) = v' = -i\alpha\phi^u(0)\exp[i\alpha(x - c\tau)].$$

From this we find

$$\eta = (\phi^u(0)/c')\exp[i\alpha(x - c\tau)] \qquad (31i)$$

in which $c' = c - U^u(0)$.

(j) The continuity in u' at interface then demands that

$$\phi'^u(0) + \phi^u(0)U'^u(0)/c' = \phi'^l(0) + \phi^l(0)U'^l(0)/c'. \qquad (31j)$$

The differential system governing the stability consists of Eqs. (27), (28), and (31a)–(31j).

V. SOLUTION OF THE DIFFERENTIAL SYSTEM

The regular perturbation technique is used to solve the eigenvalue problem for long waves ($\alpha \ll 1$). From substituting Eqs. (28) into Eqs. (27) and (31a)–(31j), we have to the first power of α,

$$\phi''''^u - i\alpha R(U^u - c)\phi''^u = 0, \quad \phi''''^l - i\alpha R\nu(U^l - c)\phi''^l/m_\eta = 0, \qquad (32)$$

$$\phi^u(1) = 0, \quad \phi^u(1) = 0, \quad \phi^l(-n) = 0, \quad \phi^l(-n) = 0, \qquad (33)$$

$$\phi^u(0) = \phi^l(0), \qquad (34)$$

$$\phi'^u(0) - \phi'^l(0) + (a_1 - a_2)\phi^u(0)/c' = 0, \qquad (35)$$

$$\phi''^u - i\alpha\Delta M(2\phi'^u U'^u + \phi''^u(U^u - c)) - m_\eta\phi''^l + i\alpha m_\eta m_\lambda\Delta M(2\phi'^l U^l + \phi''^l(U^l - c)) = -i\alpha RV_0^2\epsilon^u\epsilon^l\phi^u(0)(\epsilon^u - \epsilon^l)/c'(\epsilon^u n + \epsilon^l), \qquad (36)$$

$$\phi'''^u(1 + i\alpha\Delta Mm_\lambda(U^l - c)) + i\alpha R(c'\phi'^u + a_1\phi^u) + i\alpha\Delta MU'^u\phi''^u - m_\eta\phi'''^l(1 + i\alpha\Delta M(U^u - c)) - i\alpha\Delta Mm_\eta m_\lambda U'^l\phi''^l + i\alpha R(\nu - 1)F^{-2}\phi^u(0)/c' = 0, \qquad (37)$$

in which

$$\Delta M = M_1^u - M_2^u, \quad m_\lambda = (M_1^l - M_2^l)/(M_1^u - M_2^u) = (\lambda_1^l - \lambda_2^l)/(\lambda_1^u - \lambda_2^u) \qquad (38)$$

and primes, expect the prime on c, indicate the derivative with respect to y. Following the approach of Yih,[15] by expanding the eigenfunctions and eigenvalue in a power series of wave number α,

$$\phi^u = \phi_0 + \alpha\phi_1 + \alpha^2\phi_2 + \cdots,$$

$$\phi^l = \chi_0 + \alpha\chi_1 + \alpha^2\chi_2 + \cdots,$$

$$c = c_0 + \alpha c_1 + \alpha^2 c_2 + \cdots,$$

where the subscripts 0, 1, and 2 refer to the zeroth, first, and second approximations.

A. The zeroth-order approximation

After some rather lengthy calculations, we obtain the solution of the zeroth-order differential system as follows:

$$\phi_0 = 1 + A_1 y + A_2 y^2 + A_3 y^3, \quad \chi_0 = 1 + B_1 y + B_2 y^2 + B_3 y^3, \qquad (39)$$

in which

$$A_1 = -(m_\eta + 3n^2 + 4n^3)/2n^2(1 + n),$$
$$B_1 = (n^3 + m_\eta(4 + 3n))/2m_\eta n(1 + n),$$
$$B_2 = (n^3 + m_\eta)/n^2 m_\eta(1 + n), \qquad (40)$$
$$B_3 = (n^2 - m_\eta)/2n^2 m_\eta(1 + n),$$
$$A_2 = m_\eta B_2, \quad A_3 = m_\eta B_3.$$

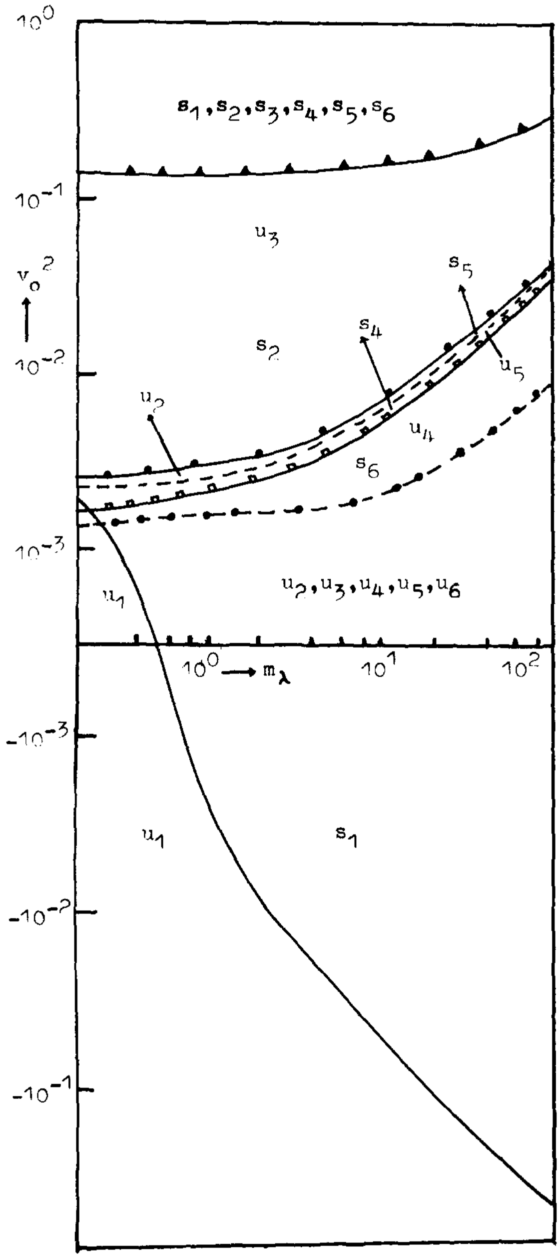

FIG. 2. A representation of the system whose particulars are $\epsilon'' = 2.284$, $\epsilon' = 80.37$, $R = 10$, $F = 1$, and $\nu = 1$. Here u denotes unstable regions and s refers to stable regions. For $m_\eta = 0.1$ and the solid curve is for $n = 0.4$, the solid curve marked with ● is for $n = 1.0$, the solid curve marked with ▲ is for $n = 10$, the solid curve marked with □ is for $n = 0.8$, the dotted curve for $n = 0.9$ marked with ● is for $n = 0.6$.

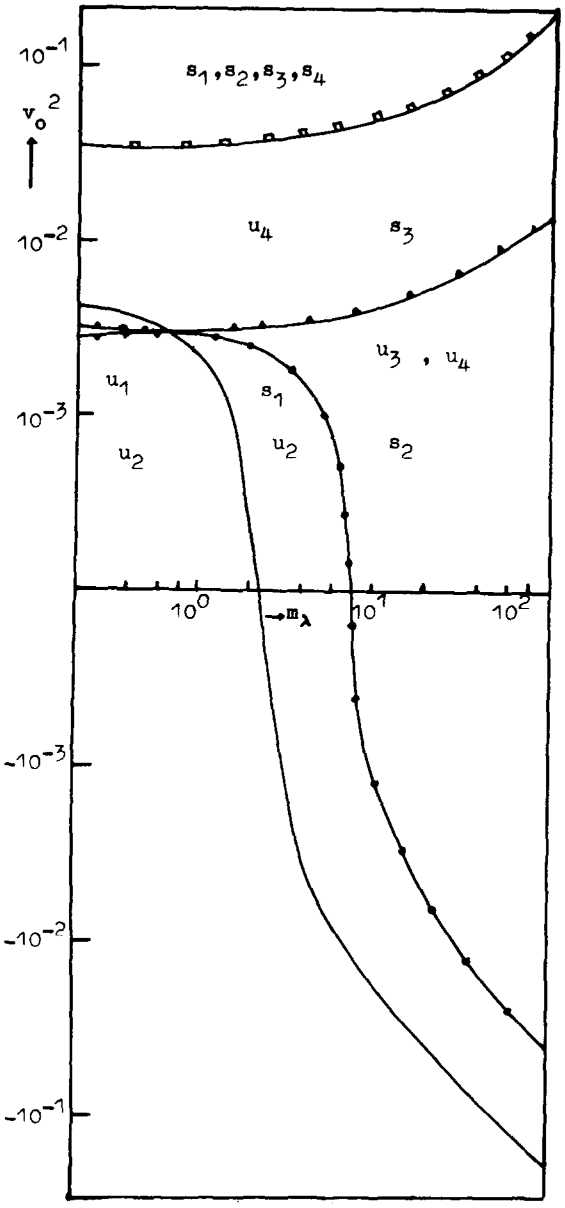

FIG. 3. The same as Fig. 2 except for $m_\eta = 0.4$, the solid curve for $n = 0.8$, the solid curve marked with ● is for $n = 0.9$, the solid curve marked with ▲ is for $n = 1.0$, the solid curve marked with □ is for $n = 10$.

The eigenvalue c_0 is determined by

$$c_0' = c_0 - b = 2n^2 m_\eta (1+n)(a_1 - a_2) \times [m_\eta^2 + 2n m_\eta (2 + 3n + 2n^2) + n^4]^{-1}. \quad (41)$$

B. The first-order approximation

Having obtained the eigenvalue c_0 and eigenfunctions ϕ_0 and χ_0, we substitute them into the equations governing the first-order approximation. After some rather lengthy calculations, the general solutions are found to be

$$\phi_1 = A_1^* y + A_2^* y^2 + A_3^* y^3 + iRH_1(y), \qquad \chi_1 = B_1^* y + B_2^* y^2 + B_3^* y^3 + iR\nu H_2(y)/m_\eta, \quad (42)$$

in which

$$H_1(y) = a_1 A_3 y^6/60 + (a_1 A_2 - 3A_3 c_0')y^5/60 - c_0' A_2 y^4/12,$$

and

$$H_2(y) = a_2 B_3 y^6/60 + (a_2 B_2 - 3B_3 c_0')y^5/60 - c_0' B_2 y^4/12.$$

The six constants of integration and hence the eigenvalue will be determined from the first-order approximation of the boundary conditions. We get

$$A_1^* + A_2^* + A_3^* + iRH_1(1) = 0,$$

$$A_1^* + 2A_2^* + 3A_3^* + iRH_1'(1) = 0,$$

$$nB_1^* - n^2 B_2^* + n^3 B_3^* - iR\nu H_2(-n)/m_\eta = 0,$$

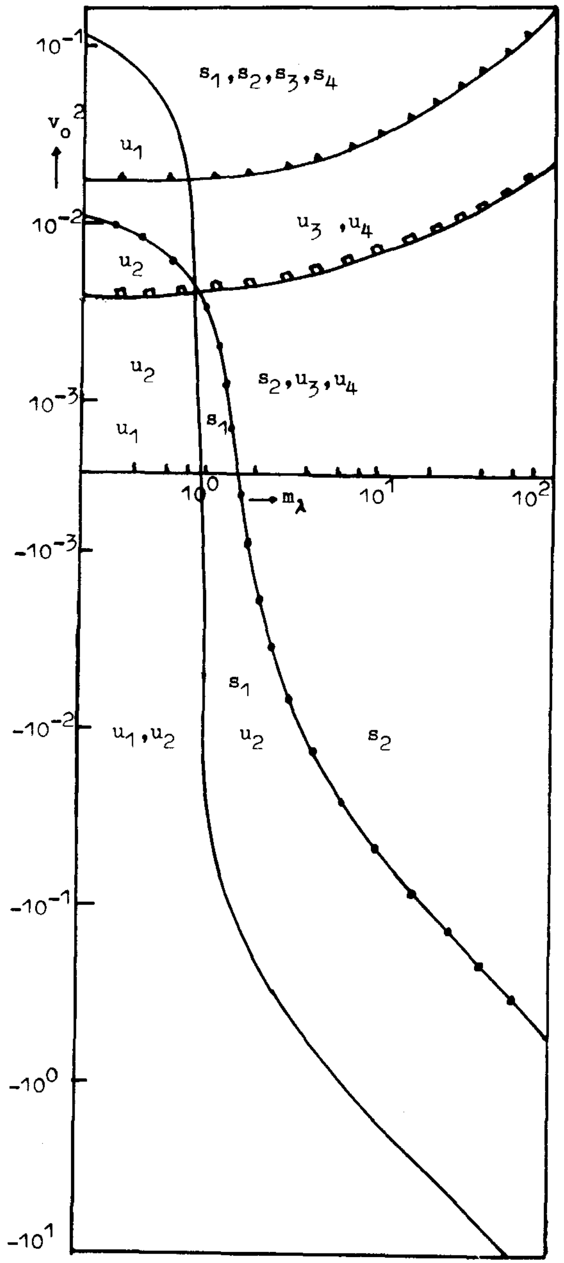

FIG. 4. The same as Fig. 2 except for $m_\eta = 0.8$ and the solid curve for $n = 0.9$, the solid curve marked with ● is for $n = 1.0$, the solid curve marked with ▲ is for $n = 10$, the solid curve marked with □ is for $n = 12$.

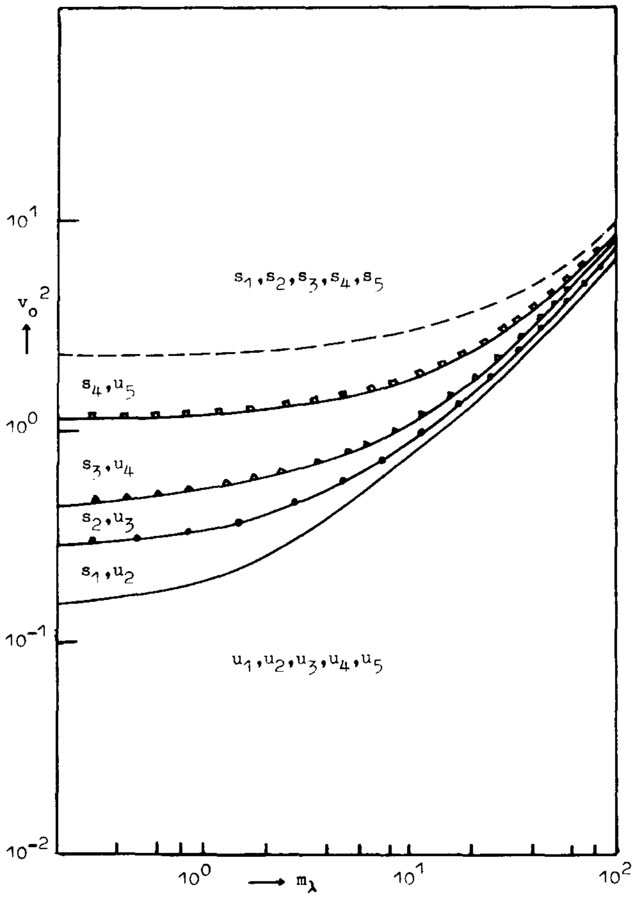

FIG. 5. The same as Fig. 2 except for $n = 0.1$ and the solid curve for $m_\eta = 1.25$, the solid curve marked with ● is for $m_\eta = 2.5$, the solid curve marked with ▲ is for $m_\eta = 0.4$, the solid curve marked with □ is for $m_\eta = 10$, and the dotted curve is for $m_\eta = 20$.

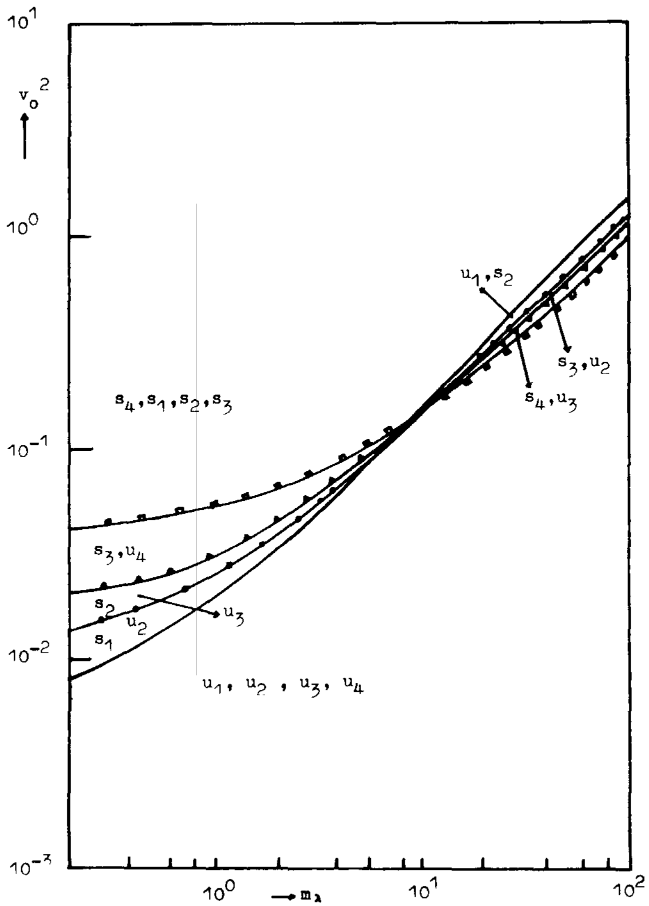

FIG. 6. The same as Fig. 2 except for $n = 0.4$, the solid curve for $m_\eta = 1.25$, the solid curve marked with ● is for $m_\eta = 2.5$, the solid curve marked with ▲ is for $m_\eta = 4.0$, and the solid curve marked with □ is for $m_\eta = 10$.

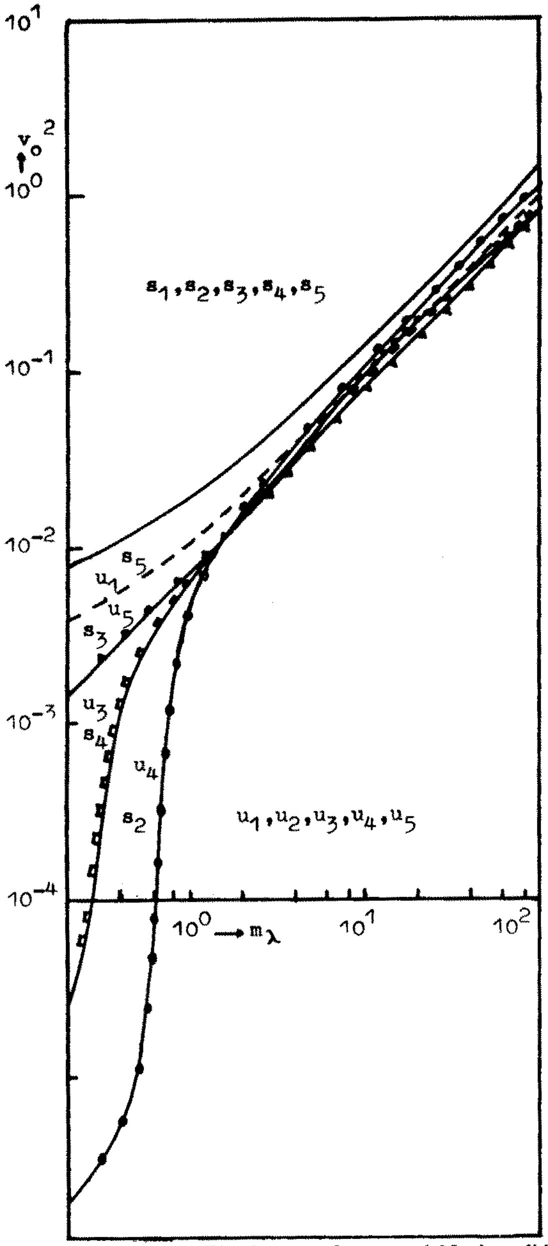

FIG. 7. The same as Fig. 2 except for $m_\eta = 1.25$, the solid curve for $n = 0.4$, the solid curve marked with ● is for $n = 1.0$, the solid curve marked with ▲ is for $n = 0.8$, the solid curve marked with □ is for $n = 0.9$, and the dotted curve is for $n = 0.6$.

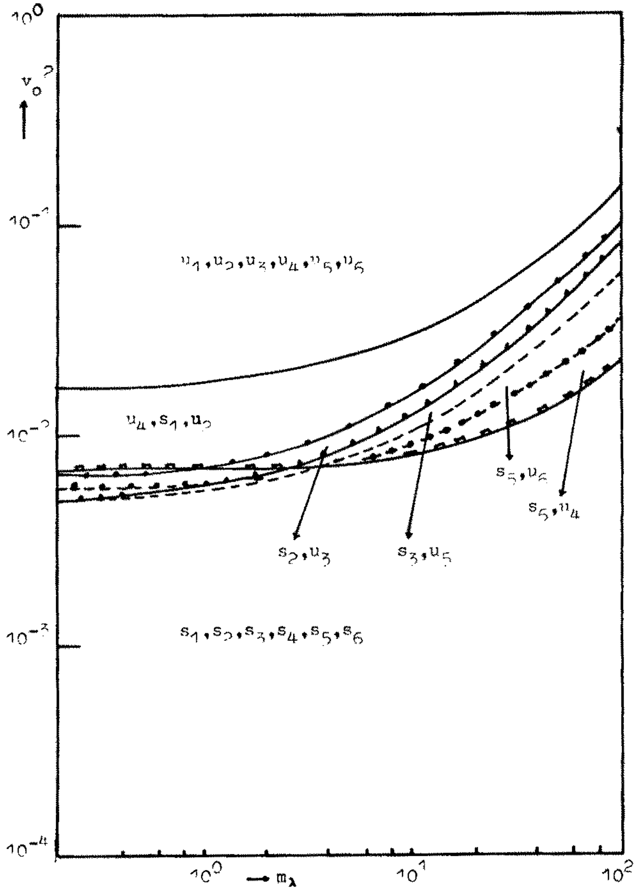

FIG. 8. The same as Fig. 2 except for $n = 10$, the solid curve for $m_\eta = 1.25$, the solid curve marked with ● is for $m_\eta = 2.5$, the solid curve marked with ▲ is for $m_\eta = 4.0$, the solid curve marked with □ is for $m_\eta = 10$, the dotted curve is for $m_\eta = 5.526$ and dotted curves marked with ● are for $m_\eta = 8$.

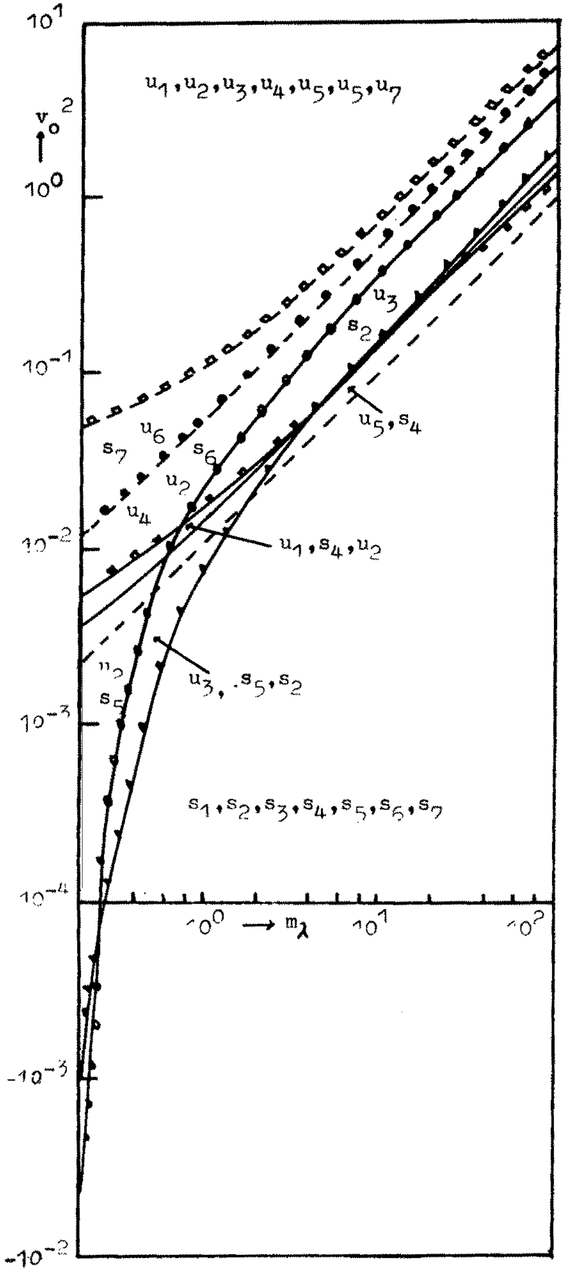

FIG. 9. The same as Fig. 2 except that the solid curve for $m_\eta = 0.4$ and $n = 0.4$, the solid curve marked with ● is for $m_\eta = 0.4$ and $n = 0.6$, the solid curve marked with ▲ is for $m_\eta = 0.8$ and $n = 0.8$, the solid curve marked with □ is for $m_\eta = 0.8$ and $n = 0.4$, the dotted curve for m_η is equal to 0.8 and $n = 0.6$, the dotted curve marked with ● is for m_η equal to 0.1 and $n = 0.1$, and the dotted curve marked with □ is for $m_\eta = 0.4$ and $n = 0.1$.

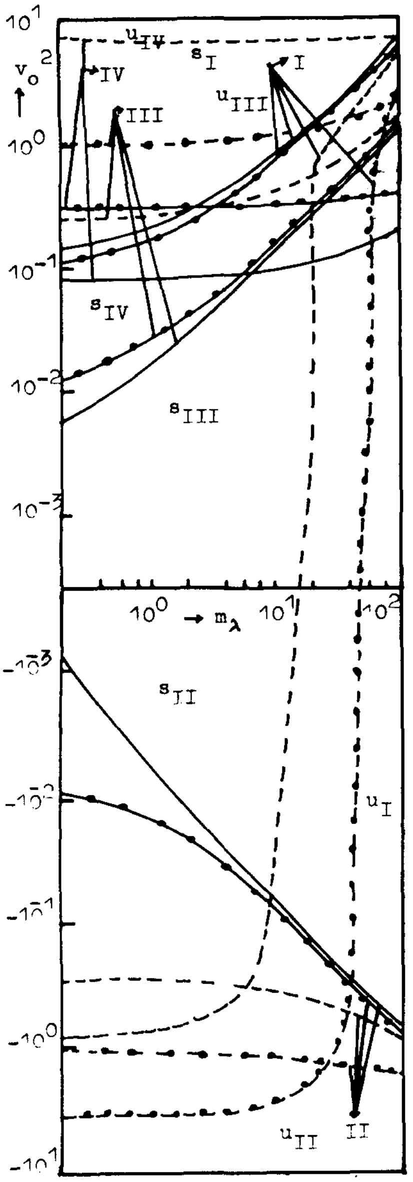

FIG. 10. For a system having $F = 0.1$ and 1.0, and $\nu = 1.0$ and 1.4, the group of curves I represent the case when $m_\eta = 1.25$ and $n = 0.1$, curves II for $m_\eta = 0.1$ and $n = 0.4$, curves III for $m_\eta = 0.4$ and $n = 0.4$, and curves IV for $m_\eta = 1.25$ and $n = 10$, where solid curves are for $F = 1.0$ and $\nu = 1.1$, the solid curves marked with ● are for $F = 1.0$ and $\nu = 1.4$, the dotted curves for $F = 0.1$ and $\nu = 1.1$, and the dotted curves marked with ● are for $F = 0.1$ and $\nu = 1.4$.

$$
\begin{aligned}
&B_1^* - 2nB_2^* + 3n^2B_3^* + iR\nu H_2'(-n)/m_\eta = 0\,, \\
&A_1^* - B_1^* + (a_2 - a_1)c_1/c_0'^2 = 0\,, \\
&A_2^* - m_\eta B_2^* - i\Delta M K_1 \\
&\qquad + iRV_0^2\epsilon^u\epsilon^l(\epsilon^u - \epsilon^l)/(\epsilon^u n + \epsilon^l) = 0\,, \\
&A_3^* - m_\eta B_3^* + iRK_2 + i\Delta M K_3 = 0\,,
\end{aligned} \tag{43}
$$

in which

$$K_1 = a_1A_1 - c_0'A_2 - m_\eta m_\lambda(a_2B_1 - c_0'B_2)\,,$$

$$K_2 = (\nu - 1)[(1/c_0'F^2) - c_0'A_1 - a_1]/6\,,$$

and

$$K_3 = \tfrac{1}{3}[a_1A_2 + 3c_0'A_3 - m_\eta m_\lambda(a_2B_2 + 3c_0'B_3)]\,.$$

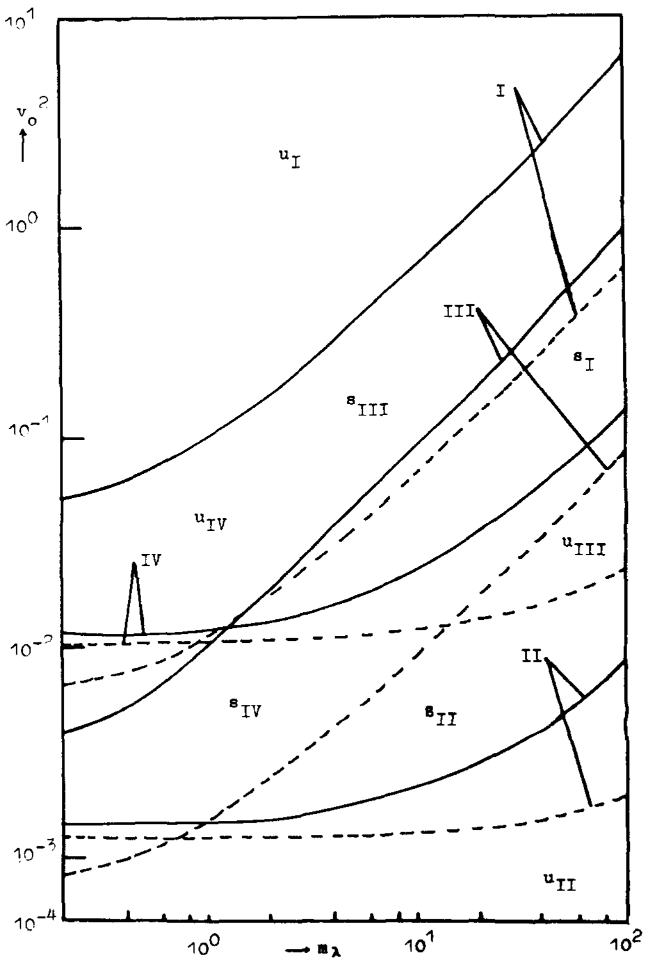

FIG. 11. For a system having $R = 10$ and 100, the group of curves I represents the case when $m_\eta = 0.4$ and $n = 0.1$, curves II for $m_\eta = 0.1$ and $n = 0.6$, curves III for $m_\eta = 1.25$ and $n = 0.6$, and curves IV for $m_\eta = 1.25$ and $n = 10$, where the solid curves are for $R = 10$ and the dotted curves are for $R = 100$.

We note that the electric field effect is felt only through the condition of continuity of shear stress at the interface in the linearized case. Also the surface tension T^* is eliminated because we take only the first power wave number. In future papers we will take into consideration the problem of nonlinear wave numbers up to the third power. This will require more complicated treatment to be solved. In this case the term of the surface tension will appear.

Separating the eigenvalue c_1 into real and imaginary parts as $c_1 = (c_1)_r + i(c_1)_i$, we obtain from Eqs. (43) the result $(c_1)_r = 0$, and

$$(c_1)_i = RJ_1 + \Delta M J_2 - RJ_1^* V_0^{*2} \tag{44}$$

in which

$$
\begin{aligned}
J_1 = {} & (c_0'^2/m_\eta(a_1 - a_2))\{m_\eta(H_1'(1) - 2H_1(1)) \\
& - 2\nu H_2(-n)/n - \nu H_2'(-n) \\
& - m_\eta K_2 + ((m_\eta - n^2)/2(1+n)) \\
& \times [H_1(1) - H_1'(1) - \nu H_2(-n)/n^2 \\
& - \nu H_2'(-n)/n + 2K_2]\}\,,
\end{aligned}
$$

$$J_1^* = c_0'^2(m_\eta - n^2)/2m_\eta(a_1 - a_2)(1+n)\,,$$

$$J_2 = (c_0'^2/m_\eta(a_1 - a_2))[(n^2 - m_\eta)K_1/2(1+n) - n(n+m_\eta)K_3/(1+n)],$$

$$V_0^{*2} = V_0^2\epsilon^u\epsilon^l(1-\epsilon^{**})/(n\epsilon^{**}+1),$$

and $\epsilon^{**} = \epsilon^u/\epsilon^l$ is the dielectric ratio.

VI. RESULTS AND DISCUSSION

The $(c_1)_i$ given by Eq. (44) is the criterion of stability. The flow is stable when $(c_1)_i$ is negative and unstable when it is positive. It consists of three parts: RJ_1 due to the viscosity, ΔMJ_2 due to the elasticity of the liquids, and $RJ_1^* V_0^{*2}$ due to the electric field. For $V_0 = 0$, Eq. (44) becomes

$$(c_1)_i = RJ_1 + \Delta MJ_2, \quad (45)$$

which is in agreement with Li's[10] result for non-Newtonian fluids, while, for $\Delta M = 0$, Eq. (45) becomes

$$(c_1)_i = RJ_1, \quad (46)$$

which reproduces Yin's[7] results for Newtonian fluids.

To see the effects of the flow phenomena, numerical calculations have been carried out, and by plotting the curves $V_0^2 = (RJ_1 + \Delta MJ_2)(n\epsilon^{**}+1)/RJ_1^*\epsilon^u\epsilon^l(1-\epsilon^{**})$ for various values of m_η (viscosity ratio), n (depth ratio), m_λ (elasticity ratio), ν (density ratio), R (Reynolds number), ϵ^{**} (dielectric ratio), and F (Froude number), we can ascertain the regions of instability and stability. We do this in Figs. 2–11 using a log–log scale in order that the effect of reciprocating the ratios can be easily observed. We discussed the stability of the system under the electric field by drawing the curves in the v_0^2-m_λ plane. The letter s stands for stable regions and u for unstable regions.

Figures 2–4 represent the cases for $m_\eta = 0.1$, 0.4, and 0.8, respectively, and different values of n where $\epsilon^{**} < 1$, $m_\eta < 1$, and $m_\eta < n^2$. The stable and unstable regions for the solid curve are shown: the solid curve marked with ●, the solid curve marked with ▲, the solid curve marked with □, and the dotted curve marked with ● are (s_1, u_1), (s_2, u_2), (s_3, u_3), (s_4, u_4), (s_5, u_5), and (s_6, u_6), respectively. We show from Fig. 2 that at $n^2 - m_\eta = 0.06$ the unstable region slightly increases, this occurs up to a value of m_λ very close to 0.4, and then the stable region increases as m_λ increases. For values of $n^2 - m_\eta = 0.26$ and higher, the stable region decreases as n increases. In Fig. 3 we observe that at $n^2 - m_\eta = 0.24$, the unstable region slightly increases up to a value of m_λ very close to 0.8, and then the stable region increases as m_λ increases, while for values of $n^2 - m_\eta = 0.6$ and higher the stable region decreases as n increases. It is clear from Fig. 4 that at $n^2 - m_\eta = 0.01$, the unstable region increases up to a value of m_λ very close to 1, and then the stable region increases as m_λ increases. But for values of $n^2 - m_\eta = 1.16$ and higher the stable region decreases as n increases.

Figures 5–7 represent the v_0^2-m_λ plane for the case where $\epsilon^{**} < 1$, $m_\eta > 1$, and $m_\eta > n^2$, but in Figs. 5 and 6 we draw curves for $n = 0.1$ and 0.4, respectively, and for different values of m_η, while in Fig. 7 we draw curves for $m_\eta = 1.25$, and different values of n. We observe from Fig. 5 that the unstable region increases as m_λ and m_η increases. In Fig. 6, we show that as m_λ increases the unstable region increases, while for m_η increases the unstable region increases up to a value of m_λ very close to 15, and then the stable region increases as m_λ increases. From Fig. 7 we can see that $m_\eta - n^2$ equals 0.89 and greater. The stable region increases as n increases, but for values of $m_\eta - n^2$ less than 0.89, the stable region decreases as n decreases to m_λ very close to 2, and then the stable region increases with m_λ.

Figure 8 illustrates the case when $\epsilon^{**} < 1$, $m_\eta > 1$, and $m_\eta < n^2$. We take the special case $n = 10$, for different values of m_η, we observe that as m_η increases not greater than 5.526, the unstable region increases. For a value of m_η greater than 5.526, the stable region increases with m_η up to a value of m_λ very close to 4, and then the unstable region increases as m_λ increases.

The case where $\epsilon^{**} < 1$, $m_\eta < 1$, and $m_\eta > n^2$ is illustrated in Fig. 9 for different values of m_η and n. We show that for $m_\eta = 0.4$ and $m_\eta - n^2 = 0.04$, the unstable region increases up to a value of m_λ very close to 0.6, and then the stable region increases as m_λ increases. For values of $m_\eta - n^2 = 0.24$ and higher the unstable region increases as n increases, while for $m_\eta = 0.8$, and $m_\eta - n^2 = 0.16$, the unstable region increases up to a value of m_λ very close to 2, and then the stable region increases as m_λ increases. For values of $m_\eta - n^2 = 0.44$ and higher the unstable region increases as n increases. Also from the same figure for $n = 0.1$ and various values of $m_\eta = 0.1$ and 0.4, respectively, we can show that the stable region increases as m_η increases. In this figure the dotted curve marked with □ divide the plane into two regions stable s_7 and unstable u_7.

Our computations are also carried out for different values of the Froude number F, and density ratio ν for the cases of $m_\eta < 1$ and $m_\eta < n^2$, $m_\eta > 1$ and $m_\eta > n^2$, $m_\eta > 1$ and $m_\eta < n^2$, and $m_\eta < 1$ and $m_\eta > n^2$. The results demonstrate that the unstable region increases while increasing the value of F, while the stable region increases while increasing the value of ν. This case is represented numerically in Fig. 10.

Figure 11 shows the variation of v_0^2 with m_λ, m_η, and n for different values of the Reynolds number R. We see that the stable region increases with the increasing value of R, this occurs for the cases having $m_\eta < 1$ and $m_\eta < n^2$, and $m_\eta > 1$ and $m_\eta > n^2$, while for the cases having $m_\eta < 1$ and $m_\eta > n^2$, and $m_\eta > 1$ and $m_\eta < n^2$, the unstable region increases with increasing the value of R.

Computations for v_0^2 are also carried out for all previous cases when ϵ^{**} changes from 0.028 42 to 35.1883. A mirror image around the m_λ axis of the curves occurs. Similar behavior occurs for all values of ϵ^{**} and its reciprocal. In order to shorten the paper all tables and some figures are not presented.

It is clear from previous results that for $\epsilon^{**} < 1$, $m_\eta \lessgtr 1$, and $m_\eta \lessgtr n^2$, the electric term is positive and the flow is stable or unstable according to whether v_0^2 is greater or less than $(RJ_1 + \Delta MJ_2)(n\epsilon^{**}+1)/RJ_1^*\epsilon^u\epsilon^l(1-\epsilon^{**})$. Also, this condition is still satisfied for $\epsilon^{**} > 1$, $m_\eta \gtrless 1$, and $m_\eta \lessgtr n^2$, while for $\epsilon^{**} < 1$, $m_\eta > 1$, and $m_\eta < n^2$, the system is stable or unstable according to whether v_0^2 is less or greater than $-(RJ_1 + \Delta MJ_2)(n\epsilon^{**}+1)/RJ_1^*\epsilon^u\epsilon^l(1-\epsilon^{**})$.

From the foregoing it can be concluded that the electric field destabilizes the flow for certain values of n, m_η, m_λ, F, ν, R, ϵ^{**}, and the electric potential v_0, and stabilizes it also, for other values of these variables.

[1]C. S. Yih, Phys. Fluids **8**, 1257 (1965).
[2]D. H. Chun and W. H. Schwabs, Phys. Fluids **11**, 5 (1968).
[3]A. S. Gupta, J. Fluid Mech. **28** (1), 17 (1967).
[4]W. Lai, Phys. Fluids **10**, 844 (1967).
[5]Fan Chun, J. Appl. Math. Phys. (ZAMP) **33**, 181 (1982).
[6]Fan Chun, J. Appl. Math. Phys. (ZAMP) **35**, 435 (1984).
[7]C. S. Yih, J. Fluid Mech. **27** (2), 337 (1967).
[8]C. E. Hickox, Phys. Fluids **14**, 251 (1971).
[9]T. W. Kao, Phys. Fluids **8**, 812 (1965).
[10]C. H. Li, Phys. Fluids **12**, 531 (1969).
[11]B. K. Shivamoggi, Astrophys. Space Sci. **84**, 477 (1982).
[12]J. G. Oldroyd, Proc. R. Soc. London Ser. A **245**, 278 (1958).
[13]H. H. Woodson and J. R. Melcher, *Electromechanical Dynamics* (Wiley, New York, 1968).
[14]J. R. Melcher, *Field Coupled Surface Waves* (MIT, Cambridge, MA, 1963).
[15]C. S. Yih, Phys. Fluids **6**, 321 (1963).

Ion acoustic solitons in a warm magnetoplasma

Kamal Kumar Ghosh
Netaji Nagar College for Women,[a] *170/436, Netaji Subhas Chandra Bose Road, Regent Park, Calcutta 700092, India*

Dipankar Ray
Department of Mathematics, Jadavpur University, Calcutta 700032, India

(Received 7 November 1985; accepted for publication 4 March 1987)

Kalita and Bujarbarua [J. Phys. A: Math. Gen. **16**, 439 (1983)] obtained a set of equations to describe the nonlinear propagation of ion acoustic waves in a warm magnetoplasma and made a numerical study of these equations for particular values of the physical parameters. In this paper a rigorous and general analytical study is presented. Some simple necessary and sufficient conditions for solitary wave solutions are derived and it is also shown that cavity solutions are not possible.

I. INTRODUCTION

For the nonlinear propagation of ion acoustic waves in a warm magnetoplasma, Kalita and Bujarbarua[1] have obtained the following evolution equation:

$$\frac{d^2F(n)}{d\eta^2} = G(n), \tag{1}$$

where

$$\begin{aligned} F(n) &= \log n + \tfrac{3}{2}\sigma n^{2/3} + M^2/2n^2, \\ G(n) &= \{ -(l_z^2/M^2)[(n-1)n \\ &\quad + \tfrac{3}{5}\sigma(n^{5/3}-1)n] + (n-1)\}, \\ \eta &= l_x X + l_z Z - Mt, \\ n &= n(\eta), \\ l_x^2 &+ l_z^2 = 1, \\ \sigma &= 5p_0/3n_0T_e. \end{aligned} \tag{2}$$

Here n is the ion density, n_0 the equilibrium ion density, p_0 the equilibrium ion pressure, T_e the electron temperature, M the velocity of nonlinear waves, and l_x, l_z are constants. The notation used here is exactly the same as that used by Kalita and Bujarbarua.[1]

The boundary conditions for localized waves are

$$\frac{dn}{d\eta} = 0 \quad \text{at } n = 1, \tag{3a}$$

and

$$n = 1 \quad \text{at } \eta = \pm\infty. \tag{3b}$$

Integrating (1) and using (3a), one gets

$$\left(\frac{dF}{dn}\right)^2\left(\frac{dn}{d\eta}\right)^2 = J(n), \tag{4a}$$

where

$$J(n) = 2\int_1^{\eta} G(n')F'(n')dn'. \tag{4b}$$

Thus to obtain localized wave solutions one has to solve Eq. (4) subject to the boundary condition (3b).

However, in view of (4a) one has to ensure that

$$J(n) \geqslant 0 \tag{I}$$

for all values of n.

Kalita and Bujarbarua explicitly evaluated the integral in (4b) and attempted to obtain the conditions under which the inequality (I) is satisfied. However, they failed to obtain general analytic results due to the complexity of the integral expression and restricted their study to a numerical integration of the equation with specific values of the constant parameters. Their results are further limited by the fact that they could check the validity of (I) at two limiting values of n and not for the intermediate region.

The purpose of this paper is to provide a general analytical solution of the above problem. It will be shown here that one can find simple necessary and sufficient conditions for the validity of (I) and hence for localized waves, by use of analytical methods without resorting to an explicit integration of the right-hand side of (4b).

II. AN ANALYTICAL STUDY

A. Necessary and sufficient conditions

From expressions for F and G given in Eq. (2), it is easy to make the following observations:

Observation 1: For $n>0$, $dF/dn = 0$ can hold at only one value of n.

Observation 2: For $n>0$, $G = 0$ can hold at most for two values of n.

Observation 3: $G(1) = 0$.

Now we establish the following theorem.

Theorem: For given values of l_x, l_z, σ, and M, the necessary and sufficient conditions for a solution of Eq. (4) that gives real n as a function of η and satisfies (3) are as follows.

Condition (i):

$$(N-n)(n-1) \geqslant 0,$$

i.e., either

$$1 \leqslant n \leqslant N \text{ or } N \leqslant n \leqslant 1,$$

for all values of n and N such that

$$J(N) = 0.$$

Condition (ii):

[a] Formerly Morning Section for Girls, Netaji Nagar College.

$$(n-1)G'(1)F'(n)>0$$

in a deleted neighborhood of $n=1$ (i.e., for $n\simeq 1$).

Condition (iii):

$$F'(1)F'(N)>0.$$

Condition (iv):

$$(N-1)F'(1)G(N)<0.$$

Proof: Conditions are necessary.

Condition (i) is necessary.

From (3b), we see that n is bounded at $\eta=\infty$. Since for our problem n is to be taken as a continuous function of η, it is bounded for finite η. Let n_1 and n_2 be, respectively, the lower and upper bounds of n. Then

$$n_1\leqslant n\leqslant n_2, \quad (5a)$$

$$\frac{dn}{d\eta}=0 \quad \text{at } n=n_1, n_2. \quad (5b)$$

Since in view of (5a), dF/dn is finite, we must have from (4a) and (5b)

$$J(n)=0 \quad \text{at } n=n_1,n_2. \quad (6)$$

By virtue of (3),

$$n_1\leqslant 1\leqslant n_2. \quad (7)$$

Then from (6) and by Rolle's theorem of differential calculus, there exist two numbers n_3, n_4 such that either

$$J'(n_3)=0, \quad n_1<n_3<1\leqslant n_2, \quad (8a)$$

or

$$J'(n_4)=0, \quad n_1\leqslant 1<n_4<n_2. \quad (8b)$$

From (4b) and (8), either

$$G(n)F'(n)=0 \quad \text{at } n=n_3, \quad (9a)$$

or

$$G(n)F'(n)=0 \quad \text{at } n=n_4. \quad (9b)$$

From observation (1) and (9), if we suppose

$$F'(n)=0 \quad \text{at } n=n_3, \quad (10)$$

then we get

$$G(n)=0 \quad \text{at } n=n_4. \quad (11)$$

From (4a) and (10), we get

$$J(n)=0 \quad \text{at } n=n_3. \quad (12)$$

From (6), (12), and by Rolle's theorem, there exist two numbers n_5 and n_6 such that

$$n_1<n_5<n_3<n_6<n_2 \quad (13a)$$

and

$$J'(n)=0 \quad \text{at } n=n_5,n_6. \quad (13b)$$

From (4b) and (13b), we get

$$G(n)F'(n)=0 \quad \text{at } n=n_5,n_6,$$

which is impossible by observations (1)–(3) and by (10) and (11). $\therefore$ $n_1<1<n_2$ is not true. $\therefore$ Either $n_1=1$ or $n_2=1$. So, we can write that either

$$1\leqslant n\leqslant N, \quad \text{where } N=n_2 \text{ if } n_1=1$$

or

$$N\leqslant n\leqslant 1, \quad \text{where } N=n_1 \text{ if } n_2=1,$$

i.e., $(N-n)(n-1)\geqslant 0$ for all values of n and N is such that $J(N)=0$, which follows from (6).

Condition (ii) is necessary.

We observe that

$$J'(n)>0 \quad (14)$$

in the deleted neighborhood of $n=1$, i.e., in the neighborhood of $n=1$ excluding this point $n=1$.

Since $G(n)$ is an analytic function of n, the Taylor series expansion of $G(n)$ around $n=1$ gives

$$G(n)=G(1)+(n-1)G'(1)+\cdots. \quad (15)$$

$\therefore$ From (4b), (14), (15), and using observation (3), we get

$$(n-1)G'(1)F'(n)>0$$

in the deleted neighborhood of $n=1$ (i.e., $n\simeq 1$).

Condition (iii) is necessary.

We shall first show that dF/dn cannot vanish in $[n_1, n_2]$. If possible, let

$$\frac{dF}{dn}=0 \quad \text{at } n=n_0\in[n_1,n_2]. \quad (16)$$

Then from (4a), (4b), (16), and (6), we get

$$J(n_0)=J(n_1)=J(n_2)=J(1)=0. \quad (17)$$

$\therefore$ From (17) and from Rolle's theorem, there exist two numbers n_7 and n_8 such that

$$n_1<n_7<n_0<n_8<n_2 \quad (18a)$$

and

$$J'(n_6)=0=J'(n_7). \quad (18b)$$

From (16) and observation (1), we conclude that $dF/dn=0$ cannot happen at $n=n_7,n_8$. Then from (4b) and (18b), we get

$$G=0 \quad \text{at } n=n_7,n_8. \quad (19)$$

From observation (2), observation (3), and Eq. (19), it follows that either $n_7=1$ or $n_8=1$.

If $n_7=1$, then from (17) and (18a) and by Rolle's theorem, we get at least one n_9 such that

$$n_1<n_9<1 \quad (20a)$$

and

$$J'(n_9)=0. \quad (20b)$$

From (4b) and (20), we see that either $G=0$ or $F'=0$ at $n=n_9$, which is impossible by observation (1), observation (2), observation (3), and Eqs. (16) and (19).

Similarly, we arrive at the same situation if $n_8=1$. So, $dF/dn\neq 0$ in $[n_1,n_2]$.

Hence $F'(1)$ and $F'(N)$ are of the same signs, i.e.,

$$F'(1)F'(N)>0.$$

Condition (iv) is necessary.

To satisfy the inequality (I) in the vicinity of $n=N$ one must have

$$J'(N)<0 \quad \text{if} \quad N>1,$$

$$J'(N)>0 \quad \text{if} \quad N<1,$$

i.e.,

$$(N-1)J'(N)<0. \quad (21)$$

From (4b), (21), and using condition (iii), we get

$$(N-1)G(N)F'(1)<0.$$

The following conditions are sufficient.

Sufficiency will be established by showing that if conditions (i)–(iv) hold then the inequality (I) holds for $(N-n)(n-1)\geqslant 0$ and boundary condition (3) is satisfied. That the inequality (I) is satisfied in the vicinity of $n=1$ and $n=N$ (i.e., for values of n sufficiently close to either 1 or N) when conditions (i)–(iv) hold can easily be established by retracing the argument given in proving that the "conditions are necessary."

To show that the inequality (I) is satisfied for the entire interval between 1 and N if conditions (i)–(iv) hold we proceed as follows: If possible, let the inequality (I) be not satisfied between 1 and N. Then there exists a number N_0 such that

$$J(N_0)<0 \quad \text{for } 1<N_0<N \text{ or } N<N_0<1.$$

But we have already proved that

$$J(n)\geqslant 0 \quad \text{for } n\approx 1 \text{ and } n\approx N.$$

Owing to the continuity of the function $J(n)$, there exist two numbers N_1, N_2 such that

$$J(N_1)=J(N_2)=J(1)=J(N)=0, \tag{22}$$

where either

$$1<N_1<N_0<N_2<N$$

or

$$N<N_2<N_0<N_1<1.$$

Then from (22) and from Rolle's theorem, there exist numbers N_3, N_4, N_5 such that

$$1<N_3<N_1<N_4<N_2<N_5<N \tag{23a}$$

or

$$N<N_5<N_2<N_4<N_1<N_3<1$$

and

$$J'(N_3)=J'(N_4)=J'(N_5)=0. \tag{23b}$$

From (4b) and (23b), we get that either

$$G=0 \quad \text{or} \quad F'=0 \quad \text{at} \quad n=N_3,N_4,N_5,$$

which is impossible by observation (2), observation (3), and condition (ii).

$\therefore$ If conditions (i)–(iv) hold then inequality (I) is satisfied for $(N-1)(N-n)<0$.

Now we establish that the boundary condition (3) is satisfied, (4a) can be written as

$$\pm\int \frac{dF/dn}{J(n)}\,dn = \pm\int d\eta, \tag{4a$'$}$$

since for $n=1$, dF/dn is finite and from (4a) we get that for $n=1$, $\eta=\pm\infty$ and hence $dn/d\eta=0$ for $n=1$.

B. Simplification of conditions

The conditions (i)–(iv) can be simplified as follows. Using the expression for $F(n)$ and $G(n)$ in Eq. (2), condition (ii) can be explicitly written as

$$M^2/l_z^2>1+\sigma>M^2 \quad \text{for } n>1 \tag{24a}$$

and

$$M^2/l_z^2<1+\sigma<M^2 \quad \text{for } n<1. \tag{24b}$$

But (24b) can hold only if $l_z^2>1$, which is impossible because $l_x^2+l_z^2=1$:

$$\therefore M^2/l_z^2>1+\sigma>M^2 \quad \text{for } n>1. \tag{25}$$

Equation (25) is equivalent to

$$G'(1)>0 \quad \text{and} \quad F'(1)>0. \tag{26}$$

Then conditions (iii) and (iv) are reduced to

$$F'(N)>0$$

and

$$(N-1)G(N)<0,$$

respectively, i.e.,

$$N^2+\sigma N^{8/3}-M^2>0$$

and

$$(N-1)\{-(l_z^2/M^2)[(N-1)N + \tfrac{3}{5}\sigma(N^{5/3}-1)N] + (N-1)\}<0. \tag{27}$$

Equation (27) can be rewritten as

$$N=K^3,$$
$$\sigma K^8+K^6>M^2, \tag{28}$$
$$\frac{M^2}{l_z^2}<K^3+\frac{3\sigma K^3(K^4+K^3+K^2+K+1)}{5(K^2+K+1)}.$$

At this stage we note that

$$\sigma K^8+K^6 \gtrless \sigma+1,$$
$$K^3+\frac{3\sigma K^3(K^4+K^3+K^2+K+1)}{5(K^2+K+1)}\gtrless 1+\sigma$$

accordingly as $K\gtrless 1$. Relations (25) and (28) can be combined to give

$$K^3+\frac{3K^3\sigma(K^4+K^3+K^2+K+1)}{5(K^2+K+1)} > \frac{M^2}{l_z^2}>1+\sigma>M^2, \tag{29}$$

for $N=K^3>1$, and $N<1$ is not possible.

III. CONCLUSION

In summary, for solutions of (1) and (2) subject to the boundary conditions (3) representing localized ion acoustic waves in a warm magnetoplasma one can get only humps (i.e., $N>1$) and cavities (i.e., $N<1$) are not possible. The conditions for humps are given by (29).

ACKNOWLEDGMENT

The authors would like to thank Professor Avijit Sen of Physical Research Laboratory, Navrangpura, Ahmedabad-380009, India, for useful suggestions regarding the presentation of the manuscript.

[1]M. K. Kalita and S. Bujarbarua J. Phys. A: Math Gen. **16**, 439 (1983).