Normally ordered Fermi operator realization of the SU_n group

Hong-yi Fan
Department of Modern Physics, China University of Science and Technology, Hefei, Anhui, People's Republic of China

(Received 13 April 1989; accepted for publication 16 August 1989)

The mapping of the SU_n group transformation in Grassmann number space into unitary normally ordered and antinormally ordered Fermi operator realizations in Hilbert space are investigated. The unitary Fermi operators are evaluated in fermion coherent state representation using the "integration within ordered product" technique for a fermionic system. Some new Fermi operator identities are thus obtained.

I. INTRODUCTION

The SU_n group has been widely applied to many fields of modern physics as a symmetry. For example, the SU_3 group is used for classifying elementary particles in strong interaction.[1] In quantum mechanics the three-dimensional isotropic harmonic oscillator possesses the SU_3 symmetry and thus is used in studying the nuclear shell model. The unitary unimodular group can also be used to develop the coherent states theory.[2] The SU_2 conserved-charge coherent state is given by Ref. 3, a construction of the SU_3 charged and hypercharged coherent states is shown in Ref. 4, and the quasicoherent state for the unitary group is obtained in Ref. 5. In this paper we shall present a so-called normally ordered Fermi operator realization of the SU_n group by using the fermion coherent state[6] and the "integration within ordered product" (IWOP) technique for a fermion system. The IWOP technique[7] for the boson system has been shown to be quite useful in studying a variety of problems in quantum mechanics and quantum optics.[8,9] We are naturally challenged to generalize the IWOP technique to the fermionic case. In Sec. II we briefly elucidate this technique and with its use we reform the completeness relation of the fermion coherent state as a normal product form. In Sec. III, using the coherent state representation, we map an SU_n transformation in Grassmann number space into the Fermi unitary operator in Hilbert space. Some identities regarding the normal product expansion of the Fermi unitary operators are thus obtained. The antinormal product expansion of the Fermi unitary operators is discussed in Sec. IV.

II. PRELIMINARIES

The fermion coherent state defined in Ref. 6 is of a qualitatively different kind than the boson coherent state[2] in that the basic label variables are Grassmann numbers. All such variables anticommute among themselves. Let α_i ($i = 1,2,...,n$) be Grassmann numbers; they are the eigenvalues of the following eigenvector equation:

$$a_i|\alpha_i\rangle = \alpha_i|\alpha_i\rangle, |\alpha_i\rangle = e^{-(1/2)\bar{\alpha}_i\alpha_i}(|0\rangle_i + |1\rangle_i\alpha_i)$$
$$= e^{-(1/2)\bar{\alpha}_i\alpha_i + a_i^\dagger\alpha_i}|0\rangle_i, \quad (2.1)$$

where a_i are Fermi annihilation operators, $a_i|0\rangle_i = 0$, $|1\rangle_i = a_i^\dagger|0\rangle_i$. Here $a_i^\dagger$ are Fermi creation operators satisfying

$$\{a_i, a_i^\dagger\} = 1, \quad a_i^2 = 0, \ a_i^{\dagger 2} = 0\,. \quad (2.2)$$

Consistency requires that α_i anticommutes with a_i and $a_i^\dagger$. The adjoint state is defined as

$$\langle\alpha_i| = {}_i\langle 0|e^{-(1/2)\bar{\alpha}_i\alpha_i + \bar{\alpha}_i a_i}, \quad \langle\alpha_i|a_i^\dagger = \langle\alpha_i|\bar{\alpha}_i\,, \quad (2.3)$$

which is not a proper mathematical adjoint because $\bar{\alpha}_i$ is independent of α_i, but as Ref. 2 mentions, one can nevertheless refer to it as such. Consistency also demands that

$$\alpha_i\bar{\alpha}_j + \bar{\alpha}_j\alpha_i = 0, \quad \bar{\alpha}_j a_i + a_i\bar{\alpha}_j = 0, \quad \bar{\alpha}_j a_i^\dagger + a_i^\dagger\bar{\alpha}_j = 0\,. \quad (2.4)$$

The fermion coherent state possesses the nonorthogonal and completeness relations

$$\langle\alpha_i'|\alpha_i\rangle = \exp[-\tfrac{1}{2}\bar{\alpha}_i\alpha_i - \tfrac{1}{2}\bar{\alpha}_i'\alpha_i' + \bar{\alpha}_i'\alpha_i]\,, \quad (2.5)$$

$$\int d\bar{\alpha}_i\, d\alpha_i|\alpha_i\rangle\langle\alpha_i| = 1\,, \quad (2.6)$$

where we used the integration rule for the Grassmann numbers

$$\int d\alpha_i = \int d\bar{\alpha}_i = 0, \quad \int d\bar{\alpha}_i\,\bar{\alpha}_i = \int d\alpha_i\,\alpha_i = 1\,. \quad (2.7)$$

In Ref. 7 we succeeded in recasting the overcompleteness relation of the boson coherent state into a normal form and thus found a variety of applications of it.[10,11] Similarly, we can put (2.6) into a normally ordered form by using

$$|0\rangle_{ii}\langle 0| = 1 - |1\rangle_{ii}\langle 1| = 1 - a_i^\dagger a_i = :e^{-a_i^\dagger a_i}: \quad (2.8)$$

and generalizing the IWOP technique to the fermion system. In sharp contrast to the boson case, where any two Bose operators commute with each other within a normal product, any two Fermi operators anticommute with each other within a normal product. For instance,

$$:a_i a_i^\dagger: = -:a_i^\dagger a_i:\,. \quad (2.9)$$

Based on property (2.9) and (2.4) we can conclude that a Grassmann number–Fermi operator (GFO) pair commutes with another GFO pair; for example,

$$:a_i\bar{\alpha}_i a_j^\dagger\alpha_j: = :a_j^\dagger\alpha_j a_i\bar{\alpha}_i:\,. \quad (2.10)$$

Because of (2.9) and (2.10) we have the following rule: A normal product of Fermi operators can be integrated or differentiated with respect to nonoperator variables; if the variable is a Grassmann number, the integration is performed according to the Berezin formula (2.7) and[12]

 0022-2488/90/020257-04$03.00

$$\int \prod_i d\bar{\alpha}_i\, d\alpha_i \exp\Big[-\sum_{i,j} \bar{\alpha}_i A_{ij} \alpha_j + \sum_i (\bar{\alpha}_i \xi_i + \bar{\xi}_i \alpha_i) \Big]$$
$$= \det A \exp\Big[\sum_{i,j} \bar{\xi}_i (A^{-1})_{ij} \xi_j \Big], \qquad (2.11)$$

where $\bar{\xi}_i$ and ξ_j are also Grassmann numbers, while A is a complex-valued matrix, as mentioned in Ref. 2. This is called the IWOP technique for a fermion system. With the use of (2.11) and the IWOP technique we are able to reform the completeness relation (2.6) as the normally ordered form

$$\int d\bar{\alpha}_i\, d\alpha_i |\alpha_i\rangle\langle\alpha_i| = \int d\bar{\alpha}_i\, d\alpha_i :\exp[-\bar{\alpha}_i\alpha_i + a_i^\dagger \alpha_i + \bar{\alpha}_i a_i - a_i^\dagger a_i]: = 1, \qquad (2.12)$$

which is essential to the following discussions.

III. MAPPING OF SU$_n$ GROUP INTO HILBERT SPACE OF FERMI OPERATORS

Consider the n-dimensional defining representation of SU$_n$, where the matrices are denoted by $[u_{ij}]$. The defining requirement of these matrices is that they are unitary and of determinant 1, e.g.,

$$\sum_i u_{ij} u_{kj}^* = \delta_{ik}, \quad \det[u_{ij}] = 1. \qquad (3.1)$$

Using the fermion coherent state we define an operator corresponding to every u:

$$U = \int \prod_i d\bar{\alpha}_i\, d\alpha_i |u\vec{\alpha}\rangle\langle\vec{\alpha}|,$$
$$\left(|\vec{\alpha}\rangle \equiv \left| \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \right\rangle = \exp\Big\{ \sum_i \Big(-\frac{1}{2}\bar{\alpha}_i\alpha_i + a_i^\dagger \alpha_i \Big) \Big\} |0\cdots 0\rangle \right), \qquad (3.2)$$

$$|u\vec{\alpha}\rangle = \left| \begin{pmatrix} \sum_i u_{1i}\alpha_i \\ \sum_i u_{2i}\alpha_i \\ \vdots \\ \sum_i u_{ni}\alpha_i \end{pmatrix} \right\rangle = \exp\Big[\sum_i \Big(-\frac{1}{2}\bar{\alpha}_i\alpha_i + \sum_j a_j^\dagger u_{ji}\alpha_i \Big) \Big] |0\cdots 0\rangle. \qquad (3.3)$$

Using the IWOP technique we integrate (3.2) to obtain

$$U = \int \prod_i d\bar{\alpha}_i\, d\alpha_i :\exp\Big\{ \sum_i \Big(-\bar{\alpha}_i\alpha_i + \sum_j a_j^\dagger u_{ji}\alpha_i + \bar{\alpha}_i a_i - a_i^\dagger a_i \Big) \Big\}: = :\exp\Big[\sum_{i,j} a_i^\dagger (u_{ij} - \delta_{ij}) a_j \Big]: \qquad (3.4)$$

or

$$U = :\exp[\vec{a}^\dagger (u - \mathbb{1})\vec{a}]:, \quad \vec{a} \equiv \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \quad \vec{a}^\dagger \equiv (a_1^\dagger a_2^\dagger \cdots a_n^\dagger), \qquad (3.5)$$

where $\mathbb{1}$ is an $n\times n$ unit matrix. Equation (3.2) indicates that for every u of the SU$_n$ group there exists a normally ordered exponential Fermi operator which is the mapping of u. Let u' be another element of the SU$_n$ group. As a result of (3.5) we have

$$U'U = :\exp[\vec{a}^\dagger (u' - \mathbb{1})\vec{a}]: :\exp[\vec{a}^\dagger (u - \mathbb{1})\vec{a}]:. \qquad (3.6)$$

On the other hand, from the original definition (3.1) of u and using the IWOP technique we obtain

$$U'U = \int \prod_i d\bar{\alpha}_i'\, d\alpha_i' |u'\vec{\alpha}'\rangle\langle\vec{\alpha}'| \int \prod_i d\bar{\alpha}_i\, d\alpha_i |u\vec{\alpha}\rangle\langle\vec{\alpha}|$$
$$= \int \prod_i d\bar{\alpha}_i'\, d\alpha_i'\, d\bar{\alpha}_i\, d\alpha_i |u'\vec{\alpha}'\rangle\langle\vec{\alpha}| \exp\Big[\sum_i \Big(-\frac{1}{2}\bar{\alpha}_i'\alpha_i' - \frac{1}{2}\bar{\alpha}_i\alpha_i + \sum_j \bar{\alpha}_i' u_{ij}\alpha_j \Big) \Big]$$
$$= \int \prod_i d\bar{\alpha}_i'\, d\alpha_i'\, d\bar{\alpha}_i\, d\alpha_i :\exp\Big\{ \sum_i \Big[-\bar{\alpha}_i'\alpha_i' - \bar{\alpha}_i\alpha_i + \sum_j (a_j^\dagger u_{ji}'\alpha_i' + \bar{\alpha}_i' u_{ij}\alpha_j) + \bar{\alpha}_i a_i - a_i^\dagger a_i \Big] \Big\}:$$
$$= \int \prod_i d\bar{\alpha}_i\, d\alpha_i :\exp\Big\{ \sum_i \Big[-\bar{\alpha}_i\alpha_i + \sum_{j,k} a_j^\dagger u_{ji}' u_{ik}\alpha_k + \bar{\alpha}_i a_i - a_i^\dagger a_i \Big] \Big\}: = :\exp[\vec{a}^\dagger (u'u - \mathbb{1})\vec{a}]:. \qquad (3.7)$$

Comparing (3.6) and (3.7) we obtain the multiplication rule for normally ordered exponential Fermi operators:

$$:\exp[\vec{a}^\dagger (u' - \mathbb{1})\vec{a}]: :\exp[\vec{a}^\dagger (u - \mathbb{1})\vec{a}]: = :\exp[\vec{a}^\dagger (u'u - \mathbb{1})\vec{a}]:. \qquad (3.8)$$

In particular, when $u^{-1} = u'$, Eq. (3.8) leads to

$$:\exp[\vec{a}^\dagger (u^{-1} - \mathbb{1})\vec{a}]: :\exp[\vec{a}^\dagger (u - \mathbb{1})\vec{a}]: = 1. \qquad (3.9)$$

It then follows that

$$:\exp[\vec{a}^\dagger (u^{-1} - \mathbb{1})\vec{a}]: = U^{-1} = \int \prod_i d\bar{\alpha}_i\, d\alpha_i |u^{-1}\vec{\alpha}\rangle\langle\vec{\alpha}|. \qquad (3.10)$$

Owing to the Grassmann integration measure $\Pi_i\, d\bar{\alpha}_i\, d\alpha_i$ being invariant under u transformation, e.g., $\det u = 1$, we can prove U defined by (3.2) as being unitary via

$$U^\dagger = \int \prod_i d\bar{\alpha}_i\, d\alpha_i |\vec{\alpha}\rangle\langle u\vec{\alpha}|$$
$$= \int \prod_i d\bar{\alpha}_i\, d\alpha_i |u^{-1}\vec{\alpha}\rangle\langle\vec{\alpha}| = U^{-1}. \qquad (3.11)$$

In particular, if the matrices u_{ij} are in the vicinity of the identity matrix δ_{ij}, e.g.,

$$u_{ij} = \delta_{ij} + \Delta u_{ij},$$

where Δu_{ij} are traceless, anti-Hermitian infinitesimal matrices, (3.4) in this case reduces to

$$U = :\exp\left[\sum_{i,j} a_i^\dagger \Delta u_{ij} a_j\right]:.$$

From (3.8) and (3.11) we conclude that the normally ordered form (3.4) makes up a realization of the SU_n group. We now prove that U can be further put into the following form:

$$U = \int \prod_i d\bar{\alpha}_i\, d\alpha_i |u\vec{\alpha}\rangle\langle\vec{\alpha}| = \exp\left[i \sum_{i,j} a_i^\dagger V_{ij} a_j\right], \quad u = e^{iV}. \qquad (3.12)$$

Proof: Let us recall the well-known result that every n-dimensional unitary matrix can be diagonalized by means of a similar transformation with a unitary matrix. Thus u can be written as

$$u = e^{iV}, \quad V^\dagger = V, \quad \mathrm{Tr}\, V = 0, \qquad (3.13)$$

where V is a traceless, Hermitian matrix characterized by $n^2 - 1$ independent real parameters. Using the operator identities

$$[AB, C] = A\{C,B\} - \{C,A\}B \qquad (3.14)$$

and

$$e^A B e^{-A} = B + [A,B] + (1/2!)[A,[A,B]] + (1/3!)[A,[A,[A,B]]] + \cdots \qquad (3.15)$$

we calculate the following transformation generated by the unitary operator $S \equiv \exp(i\Sigma_{i,j} a_i^\dagger V_{ij} a_j)$:

$$S a_i^\dagger S^{-1} = \sum_j (e^{iV})_{ji} a_j^\dagger. \qquad (3.16)$$

By noticing

$$S|0\cdots0\rangle = |0\cdots0\rangle \qquad (3.17)$$

and the fermion coherent state's completeness relation (2.10), as well as (3.16) and (3.17), we obtain

$$S = \int \prod_i d\bar{\alpha}_i\, d\alpha_i\, S|\vec{\alpha}\rangle\langle\vec{\alpha}| = \int \prod_i d\bar{\alpha}_i\, d\alpha_i\, S \times \exp\left[\sum_i a_i^\dagger \alpha_i\right] S^{-1}|0\cdots0\rangle\langle\vec{\alpha}| \exp\left[-\frac{1}{2}\sum_i \bar{\alpha}_i \alpha_i\right]$$
$$= \int \prod_i d\bar{\alpha}_i\, d\alpha_i :\exp\left[\sum_i \left(-\bar{\alpha}_i\alpha_i + \sum_j a_j^\dagger (e^{iV})_{ji} \alpha_i + \bar{\alpha}_i a_i - a_i^\dagger a_i\right)\right]: = :\exp\left[\sum_{i,j} a_i^\dagger (e^{iV} - 1)_{ij} a_j\right]:, \qquad (3.18)$$

which is a new Fermi operator identity. Comparing (3.18) with (3.4) we obtain (3.12). As an example of the new formula (3.18), we consider the generators of the SU_2 group ($n = 2$),

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad (3.19)$$

and the following Fermi operators:

$$T_\beta = \tfrac{1}{2}(a_1^\dagger a_2^\dagger)\sigma_\beta \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \quad \beta = (x,y,z). \qquad (3.20)$$

It is easily seen from (3.19) and (3.20) that T_x, T_y, and T_z satisfy SU_2 algebra. Now the operator $\exp(i\psi T_y)$, as a consequence of (3.18), is expressed as the following normal product form:

$$e^{i\psi T_y} = :\exp\left\{(a_1^\dagger a_2^\dagger)\begin{pmatrix} \cos(\psi/2) - 1 & \sin(\psi/2) \\ -\sin(\psi/2) & \cos(\psi/2) - 1 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}\right\}:$$
$$= \int \prod_{i=1}^{2} d\bar{\alpha}_i\, d\alpha_i \left| e^{(i/2)\psi\sigma_y} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \right\rangle \left\langle \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \right|, \qquad (3.21)$$

where we have used

$$e^{(i/2)\sigma_y \psi} = \begin{pmatrix} \cos(\psi/2) & \sin(\psi/2) \\ -\sin(\psi/2) & \cos(\psi/2) \end{pmatrix}.$$

To confirm the validity of (3.21), we first expand $\exp(i\psi T_y)$ using its normal product exponential form since any two Fermi operators anticommute with each other within : : ; this processing can be easily carried out and the power series terminates at the third term, e.g.,

$$e^{i\psi T_y} = :\sum_{n=0}^{\infty} \frac{1}{n!}\left[(a_1^\dagger a_1 + a_2^\dagger a_2)\left(\cos\frac{\psi}{2} - 1\right) + \sin\frac{\psi}{2}(a_1^\dagger a_2 - a_2^\dagger a_1)\right]^n:$$
$$= 1 + (a_1^\dagger a_1 + a_2^\dagger a_2)\left(\cos\frac{\psi}{2} - 1\right) + \sin\frac{\psi}{2}(a_1^\dagger a_2 - a_2^\dagger a_1)$$
$$+ \frac{1}{2!}:\left[(a_1^\dagger a_1 + a_2^\dagger a_2)^2\left(\cos\frac{\psi}{2} - 1\right)^2 + \sin^2\frac{\psi}{2}(a_1^\dagger a_2 - a_2^\dagger a_1)^2\right]:$$
$$= 1 + (a_1^\dagger a_2 - a_2^\dagger a_1)\sin\frac{\psi}{2} + (a_1^\dagger a_1 a_2 a_2^\dagger + a_2^\dagger a_2 a_1 a_1^\dagger)\left(\cos\frac{\psi}{2} - 1\right). \qquad (3.22)$$

On the other hand, using the anticommutative relation (2.2) we can also expand $\exp(i\psi T_y) = \Sigma_{n=0}^{\infty}(1/n!)(i\psi T_y)^n$ to obtain the same result as (3.22). As a consequence of (3.22) we have

$$e^{i\psi T_y} a_1 e^{-i\psi T_y} = a_1 \cos\frac{\psi}{2} - a_2 \sin\frac{\psi}{2},$$
$$e^{i\psi T_y} a_2 e^{-i\psi T_y} = a_2 \cos\frac{\psi}{2} + a_1 \sin\frac{\psi}{2}. \qquad (3.23)$$

Similarly, using (3.20) and (3.12) we have

$$e^{i\psi T_x} = \int \prod_{i=1}^{2} d\bar{\alpha}_i\, d\alpha_i \left| e^{(i/2)\psi\sigma_x} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \right\rangle \left\langle \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \right|$$
$$= :\exp\left\{ (a_1^\dagger a_2^\dagger) \begin{pmatrix} \cos(\psi/2) - 1 & i\sin(\psi/2) \\ i\sin(\psi/2) & \cos(\psi/2) - 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \right\} :. \qquad (3.24)$$

The expansion of the rhs of (3.24) can be readily done, e.g.,

$$e^{i\psi T_x} = 1 + i(a_1^\dagger a_2 + a_2^\dagger a_1)\sin(\psi/2) + (a_1^\dagger a_1 a_2 a_2^\dagger + a_2^\dagger a_2 a_1 a_1^\dagger)[\cos(\psi/2) - 1]. \qquad (3.25)$$

It is worthwhile pointing out that for $n \geqslant 3$, as far as the expansion of $U = \exp(i\Sigma_{i,j} a_i^\dagger V_{ij} a_j)$ as a power series is concerned, the exponential normal product form (3.4) of U is very convenient and useful because any two Fermi operators anticommute with each other within a normally ordered product and one need not pay attention to the anticommuting rule (2.2) in one's calculation of the expansion.

IV. ANTINORMAL EXPANSION OF U

In this section we show that the unitary operator U has the following antinormal product form:

$$U = \exp\left[i \sum_{i,j} a_i^\dagger V_{ij} a_j \right] = \vdots\, \exp\left[\sum_{i,j} a_i^\dagger (\mathbb{1} - e^{-iV})_{ij} a_j \right] \vdots\,, \qquad (4.1)$$

where $\vdots\ \vdots$ denotes the antinormal product.

Proof: From (2.2), (2.3), and (2.6) we know that an antinormally ordered Fermi operator can be expressed as

$$\vdots\, G(a_1, a_2, ..., a_n; a_1^\dagger, a_2^\dagger, ..., a_n^\dagger) \vdots = \int \prod_i d\bar{\alpha}_i\, d\alpha_i\, G(\alpha_1, \alpha_2, ..., \alpha_n; \bar{\alpha}_1, \bar{\alpha}_2, ..., \bar{\alpha}_n) |\vec{\alpha}\rangle\langle\vec{\alpha}|. \qquad (4.2)$$

Thus the rhs of (4.1) can be put into the following form:

$$\text{rhs of (4.1)} = \int \prod_i d\bar{\alpha}_i\, d\alpha_i \exp\left\{ \sum_{i,j} \bar{\alpha}_i [\delta_{ij} - (e^{-iV})_{ij}] \alpha_j \right\} |\vec{\alpha}\rangle\langle\vec{\alpha}|$$
$$= \int \prod_i d\bar{\alpha}_i\, d\alpha_i : \exp\left\{ \sum_i \left[-\sum_j \bar{\alpha}_i (e^{-iV})_{ij} \alpha_j + a_i^\dagger \alpha_i + \bar{\alpha}_i a_i - a_i^\dagger a_i \right] \right\} :$$
$$= \det(e^{-iV}) : \exp\left\{ \sum_{i,j} a_i^\dagger [(e^{iV})_{ij} - \delta_{ij}] a_j \right\} : = (3.4), \qquad (4.3)$$

which completes the proof. Also, (4.1) is a new Fermi operator identity. From (4.1) and (3.8) it follows that

$$\vdots\, \exp[\vec{a}^\dagger(\mathbb{1} - u'^{-1})\vec{a}] \,\vdots\, \vdots\, \exp[\vec{a}^\dagger(\mathbb{1} - u^{-1})\vec{a}] \,\vdots = \vdots\, \exp[\vec{a}^\dagger(\mathbb{1} - u^{-1}u'^{-1})\vec{a}] \,\vdots\,, \qquad (4.4)$$

which makes up the antinormally ordered Fermi operator realization of SU_n.

In summary, by generalizing the IWOP technique to the fermion system and exploiting the properties of the fermion coherent state we can see that the transformation u of the group SU_n in Grassmann number space is mapped to the normally ordered unitary operator U in one-to-one correspondence fashion, which makes up a normally ordered Fermi operator representation of the group SU_n (a normally ordered Bose operator representation of SU_n is given in Ref. 13). In addition, the IWOP technique makes the fermion coherent state more useful in obtaining new Fermi operator identities.

[1]See, e.g., M. Gell-Mann and Y. Ne'eman, *The Eightfold Way* (Benjamin, New York, 1964).
[2]J. R. Klauder and B.-S. Skargerstam, *Coherent States* (World Scientific, Singapore, 1985).
[3]K.-E. Eriksson and B.-S. Skargerstam, J. Phys. A: Math. Gen. **12**, 2175 (1979); J. Phys. A: Math. Gen. **14**, 545 (1981).
[4]Hong-yi Fan and Tu-nan Ruan, Commun. Theor. Phys. (Beijing) **2**, 1405 (1983).
[5]B.-S. Skargerstam, J. Phys. A: Math. Gen. **18**, 1 (1985).
[6]Y. Ohnuki and T. Kashiwa, Prog. Theor. Phys. **60**, 548 (1978).
[7]Hong-yi Fan and Tu-nan Ruan, Sci. Sin. A **27**, 392 (1984).
[8]Hong-yi Fan, H. R. Zaidi, and J. R. Klauder, Phys. Rev. D **35**, 1831 (1987); Hong-yi Fan and H. R. Zaidi, Phys. Rev. A **37**, 2985 (1988); Hong-yi Fan and J. VanderLinde, Phys. Rev. A **39**, 1552 (1989).
[9]Hong-yi Fan, Phys. Lett. A **126**, 145, 150 (1987); Phys. Lett. A **127**, 403 (1988); Phys. Lett. A **129**, 4 (1988); Hong-yi Fan and H. R. Zaidi, Phys. Lett. A **124**, 303 (1987).
[10]Hong-yi Fan, J. Math. Phys. **30**, 1273 (1989).
[11]Hong-yi Fan and J. R. Klauder, J. Phys. A: Math. Gen. **21**, L725 (1988).
[12]F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).
[13]Hong-yi Fan, Commun. Theor. Phys. **6**, 377 (1986); Commun. Theor. Phys. **7**, 125 (1987).

On continuity of mean values of unbounded observables

Jan Rusinek
Institute of Applied Mathematics and Mechanics, University of Warsaw, Pałac Kultury i Nauki, IX p., 00-901 Warsaw, Poland

(Received 20 April 1989; accepted for publication 26 July 1989)

An unbounded observable whose mean value at some state changes discontinuously under the action of a *-weakly continuous one-parameter symmetry group is exhibited. Also, a sufficient condition for differentiability of a mean value of a not necessarily bounded observable that evolves under the action of a *-weakly continuous one-parameter automorphism group is given.

I. INTRODUCTION

The main objective of the present paper is to exhibit a C^*-algebra $\mathfrak{A}$ generated, as a C^*-algebra, by two families $(U(t))_{t\in\mathbf{R}}$ and $(V(t))_{t\in\mathbf{R}}$ of unitaries, and a state ω over $\mathfrak{A}$ such that if $(\mathfrak{H}_\omega, \pi_\omega, \Omega_\omega)$ is the GNS representation of $\mathfrak{A}$ associated with ω, where $\mathfrak{H}_\omega$ is a Hilbert space, π_ω is a representation of $\mathfrak{A}$ in $\mathfrak{H}_\omega$, and $\Omega_\omega\in\mathfrak{H}_\omega$ is a cyclic vector for π_ω such that

$$\omega(A) = (\pi_\omega(A)\Omega_\omega,\Omega_\omega) \quad (A\in\mathfrak{A}),$$

then $(\pi_\omega(U(t)))_{t\in\mathbf{R}}$ and $(\pi_\omega(V(t)))_{t\in\mathbf{R}}$ are one-parameter strongly continuous unitary groups in $\mathfrak{H}_\omega$ with corresponding infinitesimal generators $i\mathscr{A}$ and $i\mathscr{B}$, and there exists a dense subspace $\mathscr{D}$ of $\mathfrak{H}_\omega$ contained in the domains of $\mathscr{A}$ and $\mathscr{B}$ such that

$$(\mathscr{A}(\mathscr{D})\cup\mathscr{B}(\mathscr{D}))\subset\mathscr{D},$$

$$(\pi_\omega(U(t))(\mathscr{D})\cup\pi_\omega(V(t))(\mathscr{D}))\subset\mathscr{D},$$

for each $t\in\mathbf{R}$, Ω_ω is in $\mathscr{D}$, and the function

$$t\rightarrow(\mathscr{B}\pi_\omega(U(t))\Omega_\omega,\pi_\omega(U(t))\Omega_\omega)$$

is bounded in no open subset of $\mathbf{R}$. It will be obvious from the construction that $\mathfrak{A}$ may be interpreted as an algebra of observables of a quantum mechanical system of one degree of freedom, the von Neumann algebra $\mathfrak{M}$ generated by $\pi_\omega(\mathfrak{A})$ as the algebra of observables at the state ω (cf. Ref. 1, p. 122), and the one-parameter group $(\tau_t)_{t\in\mathbf{R}}$ of *-automorphisms of $\mathfrak{M}$, defined by

$$\tau_t(A) = \pi_\omega(U(-t))A\pi_\omega(U(t)) \quad (A\in\mathfrak{M},\quad t\in\mathbf{R})$$

as a *-weakly continuous symmetry group of the system. Then $\mathscr{B}$, being affiliated with $\mathfrak{M}$, will be an unbounded observable whose mean value at ω evolves unboundedly and hence discontinuously under the action of $(\tau_t)_{t\in\mathbf{R}}$. In connection with this construction a problem arises as follows. Let $(U(t))_{t\in\mathbf{R}}$ and $(V(t))_{t\in\mathbf{R}}$ be two one-parameter strongly continuous unitary groups on a Hilbert space H with corresponding infinitesimal generators iA and iB. Suppose that there exists a dense subspace D of H contained in the domains of A and B such that $(A(D)\cup B(D))\subset D$ and $(U(t)(D)\cup V(t)(D))\subset$ D, for each $t\in\mathbf{R}$. Suppose, moreover, that all the functions $t\rightarrow(BU(t)\varphi,\, U(t)\varphi)$ $(\varphi\in D)$ are locally bounded. Are these functions then actually continuous? While we may not be able to resolve that problem, we shall present a result suggesting an answer in the affirmative.

II. THE CONSTRUCTION

The construction to follow is a modification of a construction from Ref. 2.

Let $C_0^\infty(\mathbf{R})$ be the space of all complex infinitely many times differentiable functions on $\mathbf{R}$ with compact support.

For each integer $n\geqslant 2$, let φ_n be a non-negative function in $C_0^\infty(\mathbf{R})$, with support in $(2^n, 2^n+3^{-n})$, such that $\varphi_n^{(k)}\leqslant 1$, for $k\leqslant n$, let, moreover,

$$\psi_n(x) = n\delta_n^{-1}\varphi_n(x-2\cdot 3^{-n}) \quad (x\in\mathbf{R}),$$

where

$$\delta_n = \int_{\mathbf{R}}\varphi_n^3(x)dx.$$

[Notice that δ_n in Ref. 2 is taken to be $(\int_{\mathbf{R}}\varphi_n^4(x)dx)^{1/2}$.] Put

$$\varphi = \sum_{n=2}^{\infty}\varphi_n \quad\text{and}\quad \psi = \sum_{n=2}^{\infty}\psi_n.$$

Let D_0 be the set of all functions of the form

$$x\rightarrow\varphi^{(k)}(x-u)\prod_{j=1}^{m}e^{it_j\psi(x-s_j)}$$

$$(s_j,t_j,u\in\mathbf{R},\quad k\in\mathbf{N}\cup,\{0\}\quad m\in\mathbf{N}),$$

and let D be the set of all functions that can be represented as sums of an element of $C_0^\infty(\mathbf{R})$ and an element of the linear space spanned by D_0. Notice that φ is an element of D.

Let $L^2(\mathbf{R})$ be the Hilbert space of all (classes of) complex square integrable functions on $\mathbf{R}$, endowed with scalar product $\langle\cdot,\cdot\rangle$ and the corresponding norm $\|\cdot\|_2$.

It is easy to see that D is a dense linear subspace of $L^2(\mathbf{R})$. Let $\mathbb{A}$ and $\mathbb{B}$ be two linear operators in $L^2(\mathbf{R})$ defined on D by setting

$$\mathbb{A}f = if' \quad\text{and}\quad \mathbb{B}f = \psi f \quad (f\in D).$$

A minor modification of an argument from Ref. 2 shows that (1) $\mathbb{A}$ and $\mathbb{B}$ are essentially self-adjoint; (2) the one-parameter strongly continuous unitary groups $(U(t))_{t\in\mathbf{R}}$ and $(V(t))_{t\in\mathbf{R}}$ generated by the closures of $i\mathbb{A}$ and $i\mathbb{B}$, respectively, take the form

$$(U(t)f)(x) = f(x-t), \quad (V(t)f)(x) = e^{it\psi(x)}f(x)$$

$$[f\in L^2(\mathbf{R}),\quad t,x\in\mathbf{R}];$$

and (3) $(\mathbb{A}(D)\cup\mathbb{B}(D))\subset D$ and $(U(t)(D)\cup V(t)(D))\subset D$, for each $t\in\mathbf{R}$.

 0022-2488/90/020261-03$03.00

Let $\mathfrak{A}$ be the C^*-subalgebra of the C^*-algebra of all linear bounded operators in $L^2(\mathbb{R})$ generated by $(U(t))_{t\in\mathbb{R}}$ and $(V(t))_{t\in\mathbb{R}}$. Let $\{p_r: r\in\mathbb{N}\}$ be a dense subset of $\mathbb{R}$. Define a state ω over $\mathfrak{A}$ by setting

$$\omega(A) = \sum_{r=1}^{\infty} 2^{-r}\|\varphi\|_2^{-2}\langle AU(p_r)\varphi, U(p_r)\varphi\rangle \quad (A\in\mathfrak{A}).$$

Let $(\mathfrak{H}_\omega, \pi_\omega, \Omega_\omega)$ be the GNS representation of $\mathfrak{A}$ associated with ω. Taking into account that, for any $A,B\in\mathfrak{A}$ and any $t\in\mathbb{R}$,

$$(\pi_\omega(U(t))\pi_\omega(A)\Omega_\omega, \pi_\omega(B)\Omega_\omega) = \omega(B^*U(t)A) = \sum_{r=1}^{\infty} 2^{-r}\|\varphi\|_2^{-2}\langle U(t)AU(p_r)\varphi, BU(p_r)\varphi\rangle,$$

and that $\{\pi_\omega(A)\Omega_\omega: A\in\mathfrak{A}\}$ is dense in $\mathfrak{H}_\omega$, it follows that the group $(\pi_\omega(U(t)))_{t\in\mathbb{R}}$ is weakly and hence strongly continuous. Similarily, the group $(\pi_\omega(V(t)))_{t\in\mathbb{R}}$ is strongly continuous. Let $\mathscr{A}$ and $\mathscr{B}$ be the infinitesimal generators of $(\pi_\omega(U(t)))_{t\in\mathbb{R}}$ and $(\pi_\omega(V(t)))_{t\in\mathbb{R}}$, respectively.

In view of (3), for any $A\in\mathfrak{A}$ and any $s,t\in\mathbb{R}$,

$$\frac{d^n}{dt^n}(\pi_\omega(V(t))\pi_\omega(U(s))\Omega_\omega, \pi_\omega(A)\Omega_\omega) = \frac{d^n}{dt^n}\omega(A^*V(t)U(s)) = \sum_{r=0}^{\infty} 2^{-r}\|\varphi\|_2^{-2}\langle \mathrm{B}^n V(t)U(p_r)\varphi, AU(p_r)\varphi\rangle.$$

Now the closedness of differentiation and the cyclicity of Ω_ω imply that the function $t\to\pi_\omega(V(t))\pi_\omega(U(s))\Omega_\omega$ is infinitely many times weakly differentiable and hence infinitely many times strongly differentiable. In particular, for each $s\in\mathbb{R}$, $\pi_\omega(U(s))\Omega_\omega$ is in the domain of $\mathscr{B}$ and, moreover,

$$(B\pi_\omega(U(s))\Omega_\omega, \pi_\omega(U(s))\Omega_\omega) = \sum_{r=1}^{\infty} 2^{-r}\|\varphi\|_2^{-2}\langle BU(s+p_r)\varphi, U(s+p_r)\varphi\rangle. \quad (2.1)$$

Given a linear operator T in a Hilbert space, let $T^{(0)}$ denote the identity operator and let $T^{(1)}$ denote T.

Proceeding along the same lines as above, we prove that for any $s_1,\ldots,s_n,t_1,\ldots,t_n\in\mathbb{R}$ and any $i_1,\ldots,i_n,j_1,\ldots,j_n,k_1,\ldots,k_n,l_1,\ldots,l_n\in\{0,1\}$, the vector

$$\mathscr{A}^{(i_1)}\pi_\omega(U(s_1))^{(j_1)}\mathscr{B}^{(k_1)}\pi_\omega(V(t_1))^{(l_1)}\times\cdots\times\mathscr{A}^{(i_n)}\pi_\omega(U(s_n))^{(j_n)}\mathscr{B}^{(k_n)}\pi_\omega(V(t_n))^{(l_n)}$$

is well defined and, for each $A\in\mathfrak{A}$,

$$(\mathscr{A}^{(i_1)}\pi_\omega(U(s_1))^{(j_1)}\mathscr{B}^{(k_1)}\pi_\omega(V(t_1))^{(l_1)}\times\cdots\times\mathscr{A}^{(i_n)}\pi_\omega(U(s_n))^{(j_n)}\mathscr{B}^{(k_n)}\pi_\omega(V(t_n))^{(l_n)}\Omega_\omega, \pi_\omega(A)\Omega_\omega)$$
$$= \sum_{r=1}^{\infty} 2^{-r}\|\varphi\|_2^{-2}\langle \mathrm{A}^{(i_1)}U(s_1)^{(j_1)}\mathrm{B}^{(k_1)}V(t_1)^{(l_1)}\times\cdots\times\mathrm{A}^{(i_n)}U(s_n)^{(j_n)}\mathrm{B}^{(k_n)}V(t_n)^{(l_n)}U(p_r)\varphi, AU(t)\varphi\rangle.$$

It is easy to see that the linear space $\mathscr{D}$ spanned by all such vectors contains Ω_ω is dense in $\mathfrak{H}_\omega$ and contained in the domains of $\mathscr{A}$ and $\mathscr{B}$, and, moreover, satisfies

$$(\mathscr{A}(\mathscr{D})\cup\mathscr{B}(\mathscr{D}))\subset\mathscr{D}$$

and

$$(\pi_\omega(U(t))(\mathscr{D})\cup\pi_\omega(V(t))(\mathscr{D}))\subset\mathscr{D},$$

for each $t\in\mathbb{R}$.

Now, turning back to (2.1), we see that, for any $r,m\in\mathbb{N}$,

$$(\mathscr{B}\pi_\omega(U(-p_r+2\cdot 3^{-m}))\Omega_\omega, \pi_\omega(U(-p_r+2\cdot 3^{-m}))\Omega_\omega) \geqslant 2^{-r}\|\varphi\|_2^{-2}\langle \mathrm{B}U(2\cdot 3^{-m})\varphi, U(2\cdot 3^{-m})\varphi\rangle.$$

Reasoning as in Ref. 2, we conclude that the right-hand side of the latter inequality does not exceed $2^{-r}\|\varphi\|_2^{-2}m$. Since $\{-p_r: r\in\mathbb{N}\}$ is dense in $\mathbb{R}$, it follows that the function

$$t\to(\mathscr{B}\pi_\omega(U(t))\Omega_\omega, \pi_\omega(U(t))\Omega_\omega)$$

is bounded in no open subset of $\mathbb{R}$.

III. A DIFFERENTIABILITY CONDITION

In this section, we shall prove the following theorem.

Theorem 3.1: Let D be a dense linear subspace of a Hilbert space H. Let A and B be two linear operators in H defined on D such that (i) $(A(D)\cup B(D))\subset D$; (ii) A is essentially self-adjoint and B is symmetric; and (iii) the closure of iA generates a one-parameter strongly continuous unitary group $(U(t))_{t\in\mathbb{R}}$ such that $U(t)(D)\subset D$, for each $t\in\mathbb{R}$, and, for each $\varphi\in D$, the function

$$(s,t)\to(BU(s)\varphi, U(t)\varphi)$$

is bounded in an open neighborhood of (O,O). Then, for any $\varphi,\psi\in D$, the function

$$(s,t)\to(BU(s)\varphi, U(t)\psi)$$

is infinitely many times differentiable in $\mathbb{R}^2$.

Proof: Let $\varphi,\psi\in D$. In view of (i), (iii), and the polarization identity, there exist $\delta>0$ and $M>0$ such that

$$|(BU(x)A\varphi, U(y)A\psi)| < M, \quad (3.1)$$

whenever $|x|<\delta$ and $|y|<\delta$. Since

$$U(w)f-f = i\int_0^w U(x)Af\,dx \quad (w\in\mathbb{R},\ f\in D),$$

it follows that, for any $s,t\in\mathbb{R}$,

$$(BU(s)\varphi - B\varphi, U(t)\psi-\psi) = -i\int_0^t (BU(s)\varphi - B\varphi, U(y)A\psi)dy = -i\int_0^t (U(s)\varphi-\varphi, BU(y)A\psi)dy$$
$$= \int_0^t\left[\int_0^s (U(x)A\varphi, BU(y)A\psi)dx\right]dy = \int_0^t\left[\int_0^s (BU(x)A\varphi, U(y)A\psi)dx\right]dy. \quad (3.2)$$

Hence, in view of (3.1), for $|s|<\delta$ and $|t|<\delta$,

$$(BU(s)\varphi - B\varphi, U(t)\psi-\psi), \quad M|s|<|t|. \quad (3.3)$$

Now, for any $s,t\in\mathbb{R}$,

$$\begin{aligned}(BU(s)\varphi,U(t)\psi) - (B\varphi,\psi) &= (BU(s)\varphi - B\varphi,U(t)\psi - \psi) + (BU(s)\varphi - B\varphi,\psi) \\ &\quad + (B\varphi,U(t)\psi - \psi) \\ &= (BU(s)\varphi - B\varphi,U(t)\psi - \psi) + (U(s)\varphi - \varphi,B\psi) \\ &\quad + (B\varphi,U(t)\psi - \psi).\end{aligned} \tag{3.4}$$

The latter identity jointly with Eq. (3.3) and the fact that $(U(t))_{t\in\mathbb{R}}$ is strongly continuous show that the function $(s,t)\to(BU(s)\varphi,U(t)\psi)$ is continuous at (O,O). Replacing φ by $U(x)\varphi$ and ψ by $U(y)\psi$ $(x,y\in\mathbb{R})$, we see that the function $(s,t)\to(BU(s)\varphi,U(t)\psi)$ is continuous everywhere.

Now, on account of (3.2), continuity of the function $(s,t)\to(BU(s)\varphi,U(t)\psi)$ ensures differentiability of the function

$$(s,t)\to(BU(s)\varphi - B\varphi,U(t)\psi - \psi).$$

Since, clearly, both functions $s\to(U(s)\varphi - \varphi,B\psi)$ and $t\to(B\psi,U(t)\psi - \psi)$ are differentiable, it follows from (3.4) that the function $(s,t)\to(BU(s)\varphi,U(t)\psi)$ is differentiable. Continuing the process, we finally arrive at the conclusion that the function $(s,t)\to(BU(s)\varphi,U(t)\psi)$ is differentiable infinitely many times.

The proof is complete.

[1]O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics 1* (Springer, New York, 1979).
[2]J. Rusinek, "Noncommuting unitary groups and local boundedness," Proc. Am. Math. Soc. **101**, 283 (1987).

Kronecker products, minuscule representations, and polynomial identities

Mircea Iosifescu and Horia Scutaru
Department of Theoretical Physics, Central Institute of Physics, P.O. Box MG-6, Bucharest, Romania

(Received 2 Februrary 1989; accepted for publication 16 August 1989)

The special role of the highest long and highest short roots in the derivation of extremum properties of the weights of irreducible representations of semisimple Lie algebras is pointed out. These properties are used to give an intrinsic and unifying reformulation, as well as a new proof of Klimyk's theorem on Kronecker products [Ukrain. Mat. Z. **18**, 19 (1966)]. This new form of Klimyk's theorem reveals the special position of the minuscule representations in Kronecker products; as an immediate consequence, the explicit formula for the Kronecker product of an arbitrary representation with a minuscule representation is obtained. Explicit expressions for the weights of the minuscule representations are given. New derivations of theorems of Dynkin [Trudy Mosk. Obsch. **1**, 39 (1952)] and Feingold [Proc. Am. Math. Soc. **70**, 109 (1978)] on Kronecker products, as well as the necessary condition for a Kronecker product to decompose in two irreducible components are obtained. The proofs are based on a theorem of Parthasarathy, Ranga Rao, and Varadarajan [Ann. Math. **85**, 383 (1967)].

I. INTRODUCTION

The irreducible representations of semisimple Lie algebras are known to satisfy specific polynomial identities.[1] There is a close connection between the identities satisfied by a representation and its Kronecker products with other representations. An immediate connection between tensorial identities and Kronecker products is given by the Wigner–Eckart theorem; we will be reminded of another connection in what follows.

The existence of polynomial identities satisfied by the generators of representations of classical Lie algebras[2] has received a simple mathematical explanation in an unpublished paper by Hannabuss.[3] Subsequent specifications by Kostant[4] and Okubo[5] have led to a method for the determination of the identities satisfied by a given irreducible representation of a semisimple Lie algebra.

The point of the method is the remark that given two irreducible L modules (Λ) and (Ω) of highest weights Λ and Ω, respectively, of a semisimple Lie algebra L of rank n, the operator

$$\mathscr{O}_{\Lambda\Omega} \equiv \sum_{i=1}^{n} e_i \otimes e^i \qquad (1.1)$$

defined on $(\Lambda) \otimes (\Omega)$ commutes with the corresponding representation of L and is expressible as a function of the second-degree Casimir operators $c_2((\Lambda)\otimes(\Omega))$, $c_2((\Lambda))$, and $c_2((\Omega))$. In formula (1.1), $\{e_i, i = 1,2,...,n\}$ denotes a basis in L and $\{e^i, i = 1,2,...n\}$ is the basis dual to $\{e_i\}$. When no confusion is possible, we shall use the same notation for an L module and the corresponding representation, i.e., denote by (Λ) either the representation with highest weight Λ or the associated L module.

Taking the matrix elements of the minimal polynomial satisfied by $\mathscr{O}_{\Lambda\Omega}$ with respect to the basis vectors of the L module (Ω) say, one obtains polynomial identities for the corresponding representation of L on (Λ).

In previous papers,[6,7] we deduced all the pairs $\{(\Lambda),(\Omega)\}$ of irreducible representations of classical Lie algebras for which this method leads to identities of the second degree: This is equivalent to finding all pairs $\{(\Lambda),(\Omega)\}$ for which the Kronecker product $(\Lambda)\otimes(\Omega)$ decomposes into two irreducible components.

The determination of these pairs was the result of successively solving the following problems.[7]

(i) Find all second-degree irreducible tensors T_{LM} in the enveloping algebra $U(L)$ of a classical semisimple Lie algebra L which vanish on irreducible representations of L, i.e., find all primitive ideals in $U(L)$ generated by second-degree polynomials. (The tensor T_{LM} transforms under the representation of L with the highest weight M.)

(ii) Having determined the tensors T_{LM}, find all the representations (Λ) (of highest weight Λ) which annulate T_{LM}.

(iii) Select among these "solutions" of the equations $T_{LM} = 0$ those pairs (Λ), (Ω) for which the Kronecker product $(\Lambda)\otimes(\Omega)$ decomposes into two irreducible components.

In this way, we obtained the following remarkable property[6,7]: The necessary condition for a Kronecker product $(\Lambda)\otimes(\Omega)$ to decompose into two irreducible components is that one of the factors be a minuscule representation. The property is rank-independent.

An irreducible representation (Λ) of highest weight Λ is called a minuscule representation if every one of its weights can be obtained by the actions on Λ of the Weyl group of L; the highest weight Λ is then called a minuscule weight.[8] Otherwise stated, a representation is minuscule if the set of its weights is a unique orbit of the Weyl group. (Other equivalent definitions are shown in Sec. II D.)

This extreme simplicity of the weight diagram of a miniscule representation (Λ) leads to special properties for the Kronecker products of (Λ) with other representations and for the polynomial identities related to these Kronecker products. One of the aims of the present article is to point out some of these properties.

The property of the minuscule representations stated above that needed the solving of the chain of problems (i)–

 0022-2488/90/020264-14$03.00

TABLE I. Elements that characterizes the Kronecker products $(\Lambda) \otimes (\Omega)$ whose Clebsch–Gordan series are of length 2. In column 4 k is an arbitrary positive integer.

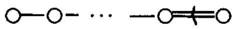
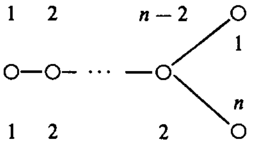
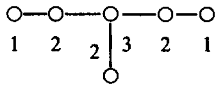
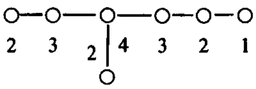
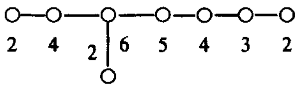
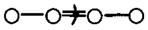
Lie algebras L	Dynkin diagrams and coefficients of the highest long roots (c_i), highest short roots (d_i), and the corresponding coroots $(c^\vee{}_i, d^\vee{}_i)$.	Minuscule weights Λ $(d^\vee{}_i = 1)$	Okubo partners Ω $k = 1,2,...$
A_n	i: 1 2 … $n-1$ n  c_i: 1 1 … 1 1 $(c_i = d_i = c^\vee{}_i = d^\vee{}_i)$	Λ_1 Λ_2 ⋮ Λ_{n-1} Λ_n	$k\Lambda_i (i = 1,2,...,n)$ $k\Lambda_1, k\Lambda_n$ ⋮ $k\Lambda_1, k\Lambda_n$ $k\Lambda_i (i = 1,2,...,n)$
B_n	i: 1 2 … $n-1$ n  c_i: 1 2 … 2 2 d_i: 1 1 … 1 1 $c^\vee{}_i$: 1 2 … 2 1 $d^\vee{}_i$: 2 2 … 2 1	Λ_n	$k\Lambda_1$
C_n	i: 1 2 … $n-1$ n  c_i: 2 2 … 2 1 d_i: 1 2 … 2 1 $c^\vee{}_i$: 1 1 … 1 1 $d^\vee{}_i$: 1 2 … 2 2	Λ_1	$k\Lambda_n$
D_n	i: 1 2 … $n-2$, $n-1$, n  c_i: 1 2 … 2; 1 ($n-1$); 1 (n) $(c_i = d_i = c^\vee_i = d^\vee_i)^1$	Λ_1 Λ_{n-1} Λ_n	$k\Lambda_{n-1}, k\Lambda_n$ $k\Lambda_1$ $k\Lambda_1$
E_6	i: 1 3 4 5 6 (2 below 4)  c_i: 1 2 3 2 1; 2 (node 2) $(c_i = d_i = c^\vee{}_i = d^\vee{}_i)$	Λ_1 Λ_6	— —
E_7	i: 1 3 4 5 6 7 (2 below 4)  c_i: 2 3 4 3 2 1; 2 (node 2) $(c_i = d_i = c^\vee{}_i = d^\vee{}_i)$	Λ_7 Λ_7	— —
E_8	i: 1 3 4 5 6 7 8 (2 below 4)  c_i: 2 4 6 5 4 3 2; 3 (node 2) $(c_i = d_i = c^\vee{}_i = d^\vee{}_i)$	—	—

TABLE I. (*Continued*).

Lie algebras L	Dynkin diagrams and coefficients of the highest long roots (c_i), highest short roots (d_i), and the corresponding coroots ($c^\vee_i, d^\vee_i$).	Minuscule weights Λ ($d^\vee_i = 1$)	Okubo partners Ω $k = 1,2,\ldots$
F_4	i: 1 2 3 4  c_i: 2 3 4 2 d_i: 1 2 3 2 $c^\vee_i$: 2 3 2 1 $d^\vee_i$: 2 4 3 2	—	—
G_2	i: 1 2  c_i: 3 2 d_i: 2 1 $c^\vee_i$: 1 2 $d^\vee_i$: 2 3	—	—

(iii), also pretends a direct proof; our initial aim was to obtain such a proof. The proof is given in Sec. V and uses a result attributed to Parthasarathy, Ranga Rao, and Varadarajan[9] which we shall call the PRV theorem.

For each minuscule weight Λ the "Okubo partner" Ω,[5,7] for which $(\Lambda) \otimes (\Omega)$ decomposes into two irreducible components, has also been determined; each minuscule weight Λ admits as Okubo partners Ω a set $\{m\Lambda_i,\ m = 1,2,\ldots\}$ of integer multiples of well-determined fundamental weights Λ_i, which are tabulated in Table I.

The existence of Okubo partners of the form $m\Lambda_i$ ($m = 1,2,\ldots$) is essential in problems of the classical limit of second-degree identities satisfied by quantum realization of semisimple Lie algebras.[6]

A number of results emerged as by-products of this study; They are perhaps as interesting as the initial question. Let us quote them briefly.

(i) We point out the special role played by the highest long and highest short roots in the derivation of extremum properties for the set of weights of an irreducible representation (Sec. III).

(ii) We obtain an intrinsic and unifying reformulation, as well as a new proof for Klimyk's theorem[10] on Kronecker products (Sec. IV). These proofs are based on the special role played by the highest long and highest short roots and therefore have a more intrinsic connection with the basic concepts of Lie algebra, avoiding the use of Weyl's formula for the characters.

(iii) This reformulation of Klimyk's theorem reveals the special position of the minuscule representations in Kronecker products; as an immediate consequence, it has led to the explicit formula for the Kronecker product of an arbitrary representation with a minuscule representation (Sec. IV B).

(iv) The highest weights of the irreducible components of the Clebsch–Gordan series of the Kronecker product between a minuscule and an arbitrary representation of highest weight Λ are those dominant weights that result from a translation with the vector Λ of the weight diagram of the minuscule representation. Therefore, we give explicit general formulas for the sets of weights of the minuscule representations for the classical Lie algebras (Appendix A) and enumerate these sets of weights for the exceptional Lie algebras (Appendix B).

(v) A new derivation of Dynkin's theorem concerning the existence of a well-defined second highest representation in the Clebsch–Gordan series of the Kronecker product of any pair of nontrivial irreducible representations[11] has been obtained (Sec. IV D), as well as a new derivation of Feingold's theorem[12] on Kronecker products for which a slight generalization has been given (Sec. IV C). Both proofs use the PRV theorem.

To conclude this introduction we would like to comment on the special role plyaed by minuscule weights in several mathematical problems in which Kronecker products and polynomial identities in $U(L)$ are involved.

For these mathematical problems, minuscule representations play a sort of minimizing role: They satisfy identities of second (minimum) degree and are factors in the nontrivial Kronecker products which decompose in Clebsch–Gordan series of minimal length.

Let us also remark that the minuscule weights classify the congruence classes of representations[13] which play an important role in Kronecker products.[14]

In addition to the algebraic aspects mentioned thus far let us also mention the geometric problems to which minuscule representations are related. For instance, in Ref. 15, Cavalli *et al.* studied the manifolds of coherent states generated from the highest weight vectors of irreducible representations of compact complex semisimple Lie groups: Their study pointed out that the manifold of coherent states generated from the highest weight vector of an irreducible representation (Λ) is a Cartan Hermitian symmetric space[16] iff(Λ) is the Okubo partner of a minuscule representation.

This property of the coherent states—the quantum states most closely related to the classical ones—is to be connected to the property of the Okubo partners of the minuscule representations to possess highest weights that are integer multiples of well-defined fundamental weights. From the identities satisfied by these representations (with highest weights of the type $m\Lambda_i$) it is possible to derive identities satisfied by Poisson bracket realizations by taking the limit for $m \to \infty$.[6]

In Ref. 17 Sakane and Takeuchi have studied the embedding of compact complex manifolds in projective spaces, which are representation spaces of semisimple Lie groups, and identified the Hermitian symmetric spaces of semisimple groups of the compact type: such symmetric spaces exist only for those semisimple Lie algebras for which minuscule representations exist; they are associated with the Okubo partners of the corresponding minuscule representations. These Hermitian symmetric spaces are proved to be described by quadrics.

Finally, let us remark that the representation spaces of minuscule representations serve as auxiliary spaces in the Yang–Baxter–Zamolodchikov–Faddeev method for the construction of completely integrable systems.[18,19]

In recent years, we assisted at a revival of the problem of the determination of the Kronecker products of irreducible representations, the story of which is too long to be traced here.[20] This revival of interest is probably explained by the feeling of the existence of deep connections between the Kronecker products of the irreducible representations of a Lie algebra, coadjoint orbits, algebraic manifolds, and primitive ideals in the enveloping algebra.

II. NOTATIONS AND DEFINITIONS

A. General notations

Let L be a complex semisimple Lie algebra of rank n; H be a Cartan subalgebra of L ($\dim H = n$); H^* be the dual space of H; R be the root system of L relative to H; ρ be the sum of the positive roots; and B be a base of simple roots $\alpha_1, \alpha_2,\dots,\alpha_n$. In addition, let $(,)$ be the nondegenerate bilinear form on H induced by the Cartan–Killing form; $\|\alpha\| \equiv \sqrt{(\alpha,\alpha)}$, be the length of the root $\alpha \in R$; $\alpha^\vee \equiv 2\alpha/(\alpha,\alpha)$, be the coroot of α; and $R^\vee \equiv \{\alpha^\vee | \alpha \in R\}$, be the coroot system of L. Let Λ^+ denote the set of dominant weights of L and $\Lambda_1, \Lambda_2,\dots,\Lambda_n$ be the fundamental weights in Λ^+ defined by

$$\langle \Lambda_i, \alpha^\vee_j \rangle \equiv 2(\Lambda_i,\alpha_j)/(\alpha_j,\alpha_j) = \delta_{ij}. \qquad (2.1)$$

We have

$$\rho = \sum_{i=1}^{n} \Lambda_i.$$

Simple roots and fundamental weights are related by

$$\alpha_i = \sum_{i=1}^{n} \langle \alpha_i, \alpha^\vee_j \rangle \Lambda_j, \qquad (2.2)$$

where $\langle \alpha_i, \alpha^\vee_j \rangle$ are the elements of the Cartan matrix of L. Any weight λ is expressible in terms of the fundamental weights

$$\lambda = \sum_{i=1}^{n} \langle \lambda, \alpha^\vee_i \rangle \Lambda_i. \qquad (2.3)$$

For any dominant weight $\Lambda \in \Lambda^+$ we denote by (Λ) the irreducible representation with highest weight Λ and by $\Pi(\Lambda)$ the set $\{\lambda\}$ of weights of representation (Λ).

B. Highest long and highest short roots

For a given semisimple Lie algebra one of the following cases occur.

(i) The system of roots divides into two sybsystems: long roots and short roots, where the lengths of the roots in each subsystem are equal. This happens for the algebras B_n, C_n, F_4, and G_2. For each of these algebras two roots are dominant weights: the highest long root α_{hl} and the highest short root α_{hs}. The other long (short) roots are obtained by acting on α_{hl} (α_{hs}) with the elements of the Weyl group. We have $\alpha_{hl} > \alpha_{hs}$. For these algebras $R^\vee \neq R$ and $\alpha^\vee_{hl}$ ($\alpha^\vee_{hs}$) are the highest short (long) roots of $R^\vee$: $\alpha^\vee_{hs} > \alpha^\vee_{hl}$.

(ii) All the roots of the algebra are of equal length, i.e., $\|\alpha_i\| = \|\alpha_j\|$ for any i,j. In this case there exists only one root which is also a dominant weight: the highest (long) root. In this case $\alpha_{hl} = \alpha_{hs}$ and $\|\alpha_{hl}\|^2 = 2$. Hence

$$\alpha^\vee_{hl} = \alpha_{hl} = \alpha_{hs} = \alpha^\vee_{hs} \qquad (2.4)$$

and $R^\vee = R$. The algebras A_n, D_n, E_6, E_7, and E_8 are of type (ii).

The roots α_{hl}, α_{hs} can be expressed as linear combinations of the simple roots α_i ($i = 1,2,\dots,n$):

$$\alpha_{hl} = \sum_{i=1}^{n} c_i \alpha_i, \quad \alpha_{hs} = \sum_{i=4}^{n} d_i \alpha_i. \qquad (2.5)$$

Similarly,

$$\alpha^\vee_{hl} = \sum_{i=1}^{n} c^\vee_i \alpha^\vee_i, \quad \alpha^\vee_{hs} = \sum_{i=4}^{n} d^\vee_i \alpha^\vee_i. \qquad (2.6)$$

The sets of coefficients c_i, d_i, $c^\vee_i$, and $d^\vee_i$ for the algebras B_n, C_n, F_4, and G_2 and $c_i = d_i = c^\vee_i = d^\vee_i$ for A_n, D_n, E_6, E_7, and E_8 are pointed out in column 2 of Table I. It is easy to prove that for any of the algebras B_n, C_n, F_4, and G_2, $c_i \geqslant d_i$ and $d^\vee_i \geqslant c^\vee_i$.

C. Dynkin portrait of a representation

Dynkin[21] proved that any weight $\lambda \in \Pi(\Lambda)$ of an irreducible finite-dimensional representation (Λ) of highest weight Λ can be written as

$$\lambda = \Lambda - \Sigma k_i \alpha_i \quad (k_i \in \mathbb{Z}^+), \qquad (2.7)$$

with $\alpha_i \in B$.

In the following we shall frequently use the "Dynkin portrait," of a representation (Λ), meaning the pattern containing all the weights of $\Pi(\Lambda)$ linked by simple roots: Two weights λ and μ ($\lambda > \mu$) are said to be linked by the simple root α_i, or to be α_i-linked, if $\lambda - \mu = \alpha_i$. This property is represented graphically by

$$\lambda \xrightarrow{\alpha_i} \mu. \qquad (2.8)$$

An "α_i string of length p" of the representation (Λ) is a subset of $\Pi(\Lambda)$ composed of $p+1$ α_i-linked weights in the Dynkin portrait of (Λ),

$$\lambda_{\max} \equiv \lambda_1 \xrightarrow{\alpha_i} \lambda_2 \xrightarrow{\alpha_i} \cdots \xrightarrow{\alpha_i} \lambda_p \xrightarrow{\alpha_i} \lambda_{p+1} \equiv \lambda_{\min}, \qquad (2.9)$$

and such that there does not exist in the Dynkin portrait of (Λ) a set of α_i-linked weights containing it. A weight λ_i of the α_i string (2.9) that is not one of its extreme elements $(\lambda_{\min}, \lambda_{\max})$ is said to be "traversed" by this α_i string.

Let λ be a weight belonging to an α_i string. The following property is well known.[8]

Proposition 2.1: Let $\Lambda \in \Lambda^+$, $\lambda \in \Pi(\Lambda)$, and $\alpha_i \in B$. Let $p_i(\lambda)$ and $q_i(\lambda)$ be the maximum non-negative integers for which

$$\lambda + p_i(\lambda)\alpha_i \in \Pi(\Lambda), \quad \lambda - q_i(\lambda)\alpha_i \in \Pi(\Lambda). \qquad (2.10)$$

Then

$$\langle \lambda, \alpha^{\vee}{}_i \rangle = q_i(\lambda) - p_i(\lambda). \qquad \square(2.11)$$

In particular, for the extreme elements of the α_i string (2.9) (of length p), the values p_i and q_i are

$$\begin{aligned} &p_i(\lambda_1) = 0, \quad q_i(\lambda_1) = p; \\ &p_i(\lambda_{p+1}) = p, \quad q_i(\lambda_{p+1}) = 0, \end{aligned} \qquad (2.12)$$

from which

$$\langle \lambda_1, \alpha_i^{\vee} \rangle = p, \quad \langle \lambda_{p+1}, \alpha_i^{\vee} \rangle = -p. \qquad (2.13)$$

Corollary 2.2: The extreme elements in an α_i string of length p have expressions of the form

$$\begin{aligned} \lambda_{\max} &= p\Lambda_i + \sum_{k \neq i} p_k \Lambda_k \quad (p_k \in \mathbb{Z}), \\ \lambda_{\min} &= -p\Lambda_i + \sum_{k \neq i} p'_k \Lambda_k \quad (p'_k \in \mathbb{Z}). \end{aligned} \qquad (2.14)$$

Corollary 2.3: The highest length of a string in the Dynkin portrait of a representation of highest weight Λ is

$$l_h = \sup_{\substack{\lambda \in \Pi(\Lambda) \\ i = 1,2,\dots,n}} \langle \lambda, \alpha^{\vee}{}_i \rangle. \qquad (2.15)$$

The Dynkin portrait of a representation (Λ) splits into layers: the kth layer contains the weights $\lambda \in \Pi(\Lambda)$ for which $\Sigma k_i = k$, with the integers k_i defined by Eq. (2.7).

The number of layers of the Dynkin portrait of representation (Λ) is equal to[11]

$$T(\Lambda) = k_{\max}(\Lambda) = 1 + \sum_{\alpha_k \in B} c_{\alpha_k}(\Lambda), \qquad (2.16)$$

where $c_{\alpha_k}(\Lambda)$ are the coefficients in the decomposition

$$\Lambda = \sum_{\alpha_k \in B} c_{\alpha_k}(\Lambda)\alpha_k. \qquad (2.17)$$

Dynkin portraits are "spindle shaped."[11] For self-contragredient representations the weights of a Dynkin portrait that are symmetric with respect to the "middle of the spindle" have opposite signs.

D. Minuscule weights and representations

A minuscule weight is a dominant weight Λ characterized by any of the following equivalent properties.[8]

(i) The dominant weight Λ is minuscule iff all the weights of the representation (Λ) [i.e., all $\lambda \in \Pi(\Lambda)$] belong to a unique orbit of the Weyl group.

(ii) The dominant weight Λ is minuscule iff all the weights of the representation (Λ) have the same length.

(iii) The dominant weight Λ is minuscule iff for any $\lambda \in \Pi(\Lambda)$ and any $\alpha \in B$

$$\langle \lambda, \alpha^{\vee} \rangle \in \{-1, 0, 1\}, \qquad (2.18)$$

i.e., for any $\lambda \in \Pi(\Lambda)$ the coefficients c_i in the development (2.3) can take only the values 0, ± 1.

(iv) The dominant weight Λ is minuscule iff

$$\langle \Lambda, \alpha^{\vee}{}_{hs} \rangle = 1. \qquad (2.19)$$

(v) The dominant weight Λ is minuscule if it is a fundamental weight Λ_i for which [cf. the notations introduced in Eqs. (2.6)]

$$d^{\vee}{}_i = 1. \qquad (2.20)$$

The set of the minuscule weights for the semisimple Lie algebras is pointed out in column 3 of Table I.

Definition (iii) and Corollary 2.3 imply that the Dynkin portrait of a minuscule representation contains only α strings of length $p = 1$. In other terms, for a minuscule representation (Λ) any weight $\lambda \in \Pi(\Lambda)$ is either the beginning or the end of an α_i string. These properties allow us to determine for any weight of a minuscule representation the strings to which they belong, i.e., for which they are either the beginning or the end. For instance, in the Dynkin portrait of representation (Λ_7) of E_7 the weight $-\Lambda_3 + \Lambda_4 - \Lambda_6 + \Lambda_7$ has the α_i links

$$\begin{array}{ccc} \alpha_3 \searrow & & \swarrow \alpha_6 \\ & -\Lambda_3 + \Lambda_4 - \Lambda_6 + \Lambda_7 . & \\ \swarrow \alpha_4 & & \searrow \alpha_7 \end{array}$$

III. INEQUALITIES FOR THE WEIGHTS OF A REPRESENTATION

In this section we point out that the highest long and highest short roots are useful in the derivation of extremum properties for the set of weights of an irreducible representation (Λ) of a semisimple Lie algebra L.

Proposition 3.1: For any weight $\lambda \in \Pi(\Lambda)$ and for any long (short) root α_j of L,

$$|\langle \lambda, \alpha^{\vee}{}_j \rangle| \leqslant \langle \Lambda, \alpha^{\vee}{}_{hl(hs)} \rangle, \qquad (3.1)$$

where $\alpha_{hl(hs)}$ denotes the highest long or the highest short root of L if α_j is a long or short root, respectively.

Proof: The difference $\Lambda - \lambda$ is a linear combination of simple roots with non-negative coefficients [Eq. (2.7)]. Observing that $\alpha^{\vee}{}_{hl}, \alpha^{\vee}{}_{hs} \in \Lambda^+$, we obtain

$$\langle \lambda, \alpha^{\vee}{}_{hl} \rangle \leqslant \langle \Lambda, \alpha^{\vee}{}_{hl} \rangle, \quad \langle \lambda, \alpha^{\vee}{}_{hs} \rangle \leqslant \langle \Lambda, \alpha^{\vee}{}_{hs} \rangle. \qquad (3.2)$$

For dominant weights $(\lambda \in \Lambda^+)$ the proof is immediate because for any long (short) root α we have [writing $\alpha^{\vee}{}_{hl(hs)} - \alpha^{\vee}$ in the form (2.7)]

$$\langle \lambda, \alpha^{\vee} \rangle \leqslant \langle \lambda, \alpha^{\vee}{}_{hl(hs)} \rangle, \qquad (3.3)$$

from which

$$0 \leqslant \langle \lambda, \alpha^\vee \rangle \leqslant \langle \Lambda, \alpha^\vee{}_{hl(hs)} \rangle. \tag{3.4}$$

Let us now assume that $\lambda \notin \Lambda^+$. Then there exists a dominant weight $\lambda_+ \in \Lambda^+$ and an element w of the Weyl group W such that

$$\lambda_+ = w\lambda \tag{3.5}$$

and Eqs. (3.2) can be written for λ_+. To complete the proof it is sufficient to prove that for $\alpha =$ long (short) root the inequality

$$|\langle \lambda, \alpha^\vee \rangle| \leqslant \langle \lambda_+, \alpha^\vee{}_{hl(hs)} \rangle \tag{3.6}$$

holds. For any such α there exists an element $w_\alpha \in W$ for which

$$w_\alpha \alpha = \alpha_{hl(hs)}. \tag{3.7}$$

Relation (3.6), which has to be proved, therefore becomes

$$|\langle w_\alpha \alpha, \alpha^\vee{}_{hl(hs)} \rangle| \leqslant \langle \lambda_+, \alpha^\vee{}_{hl(hs)} \rangle. \tag{3.8}$$

If $w_\alpha \lambda \in \Lambda^+$ the property $\lambda_+ \equiv w\lambda \in \Lambda^+$ requires the equality sign in (3.8) and $w_\alpha = w$.

The nontrivial case to be considered is

$$w_\alpha \lambda \notin \Lambda^+ \tag{3.9}$$

and we have to prove that for those $w' \in W$ for which

$$w'\lambda_+ = w_\alpha \lambda \tag{3.10}$$

we have

$$|\langle w'\lambda_+, \alpha^\vee{}_{hl(hs)} \rangle| \leqslant \langle \lambda_+, \alpha^\vee{}_{hl(hs)} \rangle \tag{3.11}$$

for any $\lambda_+ \in \Lambda^+$. If

$$\langle w'\lambda_+, \alpha^\vee{}_{hl(hs)} \rangle \geqslant 0, \tag{3.12}$$

then from

$$\lambda_+ - w'\lambda_+ = \Sigma k_{\alpha_i} \alpha_i, \quad (k_{\alpha_i} \in \mathbb{Z}^+, \quad \alpha_i \in B), \tag{3.13}$$

and $\alpha^\vee{}_{hl(hs)} \in \Lambda^+$ the inequality (3.11) follows.

Assume now

$$\langle w'\lambda_+, \alpha^\vee{}_{hl(hs)} \rangle < 0. \tag{3.14}$$

Then

$$\begin{aligned} |\langle w'\lambda_+, \alpha^\vee{}_{hl(hs)} \rangle| &= -\langle w'\lambda_+, \alpha^\vee{}_{hl(hs)} \rangle \\ &= \langle s_{\alpha_{hl(hs)}} w'\lambda_+, \alpha^\vee{}_{hl(hs)} \rangle \\ &\leqslant \langle \lambda_+, \alpha^\vee{}_{hl(hs)} \rangle, \end{aligned} \tag{3.15}$$

where $s_{\alpha_{hl(hs)}}$ is the Weyl reflection associated with the root $\alpha_{hl(hs)}$. The proof is complete. □

Comment: We have observed that Proposition 3.1 can also be obtained by combining Lemmas 2, 4, and 5 in Ref. 12. It has also been rediscovered in Ref. 17 (Lemma 3.6).

The present formulation seems to be more transparent than the previous ones; we have included our proof for the sake of clarity and self-consistency.

Remark 1: The highest bound $\langle \Lambda, \alpha^\vee{}_{hl(hs)} \rangle$ for $|\langle \lambda, \alpha^\vee{}_i \rangle|$ is effectively attained, i.e., for any representation (Λ) there exists a weight $\lambda \in \Pi(\Lambda)$ for which the equality in Eq. (3.1) holds.

Remark 2: For the fundamental representations Λ_i of any of the classical Lie algebras A_n, B_n, C_n, and D_n we have $\langle \Lambda_i, \alpha^\vee{}_{hl} \rangle \leqslant \langle \Lambda_i, \alpha^\vee{}_{hs} \rangle \leqslant 2$.

Proposition 3.2: Let $\lambda \in \Pi(\Lambda)$ and let α_i be a simple root. Let $\lambda + p_i(\lambda)\alpha_i$ and $\lambda - q_i(\lambda)\alpha_i$ be the maximum and the minimum elements, respectively, of an α_i string through λ. The non-negative integers $p_i(\lambda)$ and $q_i(\lambda)$ satisfy the relations

$$p_i(\lambda) + q_i(\lambda) \leqslant \langle \Lambda, \alpha^\vee{}_{hl(hs)} \rangle \tag{3.16}$$

if α_i is a long (short) root and $\alpha_{hl(hs)}$ has the same meaning as given previously.

Proof: Let us observe that

$$\lambda + p_i(\lambda)\alpha_i \in \Pi(\Lambda), \quad \lambda - q_i(\lambda)\alpha_i \in \Pi(\Lambda). \tag{3.17}$$

Therefore, in Eq. (3.1) we can consider $\lambda + p_i(\lambda)\alpha_i$ instead of λ. Recalling that $\langle \alpha_i, \alpha^\vee{}_i \rangle = 2$, $p_i(\lambda) \geqslant 0$, $q_i(\lambda) \geqslant 0$, and taking into account Eq. (2.11), we obtain

$$\begin{aligned} \langle \Lambda, \alpha^\vee{}_{hl(hs)} \rangle &\geqslant |\langle \lambda, \alpha^\vee{}_i \rangle + 2p_i(\lambda)| \\ &= \langle \lambda, \alpha^\vee{}_i \rangle + 2p_i(\lambda) = p_i(\lambda) + q_i(\lambda). \end{aligned} \tag{3.18}$$

□

Corollary 3.3: For any $\lambda \in \Pi(\Lambda)$ and any long (short) simple root α_i the following inequalitites hold:

$$p_i(\lambda) \leqslant \langle \Lambda, \alpha^\vee{}_{hl(hs)} \rangle, \quad q_i(\lambda) \leqslant \langle \Lambda, \alpha^\vee{}_{hl(hs)} \rangle. \tag{3.19}$$

Proposition 3.4 gives sharper higher bounds for $p_i(\lambda)$ and $q_i(\lambda)$.

Proposition 3.4: Let (Λ) be an irreducible representation of a semisimple Lie algebra L. For any weight $\lambda \in \Pi(\Lambda)$ and for any long (short) root α_i of L the inequalities

$$p_i(\lambda) \leqslant [\tfrac{1}{2}(\langle \Lambda, \alpha^\vee{}_{hl(hs)} \rangle - \langle \lambda, \alpha^\vee{}_i \rangle)], \tag{3.20}$$

$$q_i(\lambda) \leqslant [\tfrac{1}{2}(\langle \Lambda, \alpha^\vee{}_{hl(hs)} \rangle + \langle \lambda, \alpha^\vee{}_i \rangle)] \tag{3.21}$$

hold, where $[x]$ denotes the largest integer with the property $x - [x] < 1$.

Proof: Equation (3.20) results directly from Eq. (3.18). Let us introduce in Eq. (3.1) $\lambda - q_i(\lambda)\alpha_i$ instead of λ. We obtain the following relations, similar to Eqs. (3.18):

$$\begin{aligned} |\langle \lambda, \alpha^\vee{}_i \rangle - 2q_i(\lambda)| &= -(p_i(\lambda) + q_i(\lambda)) \\ &= -(\langle \lambda, \alpha^\vee{}_i \rangle - 2q_i(\lambda)) \leqslant \langle \Lambda, \alpha^\vee{}_{hl(hs)} \rangle, \end{aligned} \tag{3.22}$$

from which Eq. (3.21) follows. □

Application: Let us observe that for any semisimple Lie algebra $\alpha^\vee{}_{hl} \leqslant \alpha^\vee{}_{hs}$. Hence, if Λ is a minuscule weight, because the expression of $\alpha^\vee{}_{hl}$ as a linear combination of simple coroots [(2.6)] has non-negative integer coefficients, we obtain from Eq. (2.19) that

$$\langle \Lambda, \alpha^\vee{}_{hl} \rangle = \langle \Lambda, \alpha^\vee{}_{hs} \rangle = 1. \tag{3.23}$$

From the property (2.18) of the minuscule weights we obtain, using Eqs. (3.20) and (3.21), that

if $\langle \lambda, \alpha^\vee{}_i \rangle = 0$, then $p_i(\lambda) = q_i(\lambda) = 0$;

if $\langle \lambda, \alpha^\vee{}_i \rangle = 1$, then $p_i(\lambda) = 0$, $q_i(\lambda) = 1$;

if $\langle \lambda, \alpha^\vee{}_i \rangle = -1$, then $p_i(\lambda) = 1$, $q_i(\lambda) = 0$.

We thus prove the remarks that conclude Sec. II D.

Proposition 3.5: Let α_i be a long (short) root of the semisimple Lie algebra L; let $\Lambda, \Omega \in \Lambda^+$ be dominant weights of L with the property

$$\langle \Omega + \rho, \alpha^\vee{}_i \rangle = \langle \Omega, \alpha^\vee{}_i \rangle + 1 \geqslant \langle \Lambda, \alpha^\vee{}_{hs(hs)} \rangle \tag{3.24}$$

and let $\lambda\in\Pi(\Lambda)$ be a weight of representation (Λ) such that

$$\langle\lambda+\Omega,\alpha^\vee{}_i\rangle\geqslant 0. \tag{3.25}$$

Then

$$p_i(\lambda)\leqslant\langle\Omega,\alpha^\vee{}_i\rangle, \tag{3.26}$$

$$q_i(\lambda)\leqslant\langle\lambda+\Omega,\alpha^\vee{}_i\rangle. \tag{3.27}$$

Proof: From Eqs. (3.24) and (3.25) we obtain

$$\langle\Omega,\alpha^\vee{}_i\rangle+\tfrac{1}{2}\geqslant\tfrac{1}{2}(\langle\Lambda,\alpha^\vee{}_{hl(hs)}\rangle-\langle\lambda,\alpha^\vee{}_i\rangle). \tag{3.28}$$

Taking the integer part of both sides of Eq. (3.28) and using Eq. (3.20) we obtain the first inequality. The second inequality results from using Eq. (2.11). □

Remark 3: The sharper result (3.20) leads to a sharper inequality than the use of inequalities (3.24) and (3.19).

IV. KRONECKER PRODUCTS

The present section points out the role of the highest long and highest short roots in several theorems concerning Kronecker products. The extremum properties for the weights of a representation, which have been derived in Sec. III, will be essential for the proofs of these theorems.

Another result that will intervene in the proofs is the PRV theorem on Kronecker products.[9] The proof of the PRV theorem does not appeal to Weyl's formula for the characters; the theorem is transparent and easy to apply.

With the aid of these instruments we shall give simple proofs to several important theorems of Klimyk,[10] Feingold[12] (in a slightly more general formulation), and Dynkin.[11]

A. The PRV theorem[9]

Let L be a semisimple Lie algebra of rank n; let H be a Cartan subalgebra of L; and, for each simple root α_i of L, let h_{α_i} be the element of H defined by the equality $\alpha_i(h_{\alpha_j})=\langle\alpha_i,\alpha^\vee{}_j\rangle$. Let $x_{\alpha_i}, y_{\alpha_i}, h_{-\alpha_i}$, $i=1,2,\dots,n$ be the generators of L satisfying the relations

$$\begin{aligned}&[h_{\alpha_i},x_{\alpha_j}]=\langle\alpha_j,\alpha^\vee{}_i\rangle x_{\alpha_j},\\ &[h_{\alpha_i},y_{\alpha_j}]=-\langle\alpha_j,\alpha^\vee{}_i\rangle y_{\alpha_j},\\ &[x_{\alpha_i},y_{\alpha_i}]=h_{\alpha_i},\quad [x_{\alpha_i},y_{\alpha_j}]=0,\quad \text{if } i\neq j.\end{aligned} \tag{4.1}$$

We shall state the PRV theorem in the formulation given in Ref. 22 and using the notations introduced in Sec. II A.

The PRV theorem: Let (Λ) and (Ω) be finite-dimensional irreducible L modules of the semisimple Lie algebra L of rank n. Let (Λ,λ) be the subspace of the weight vectors of weight λ in (Λ) and let (Ω,Λ,λ) be the subspace of (Λ,λ) defined by

$$(\Omega,\Lambda,\lambda)\equiv\{v\in(\Lambda,\lambda)\,|\,x_i^{\langle\Omega+\rho,\alpha_i^\vee\rangle}v=0,\quad i=1,2,\dots,n\}. \tag{4.2}$$

Then the Clebsch–Gordan reduction of the tensor product module $(\Lambda)\otimes(\Omega)$ may be written as

$$(\Lambda)\otimes(\Omega)=\bigoplus_{\substack{\lambda\in\Pi(\Lambda)\\ \lambda+\Omega\in\Lambda^+}}(\dim(\Omega,\Lambda,\lambda))(\lambda+\Omega)\qquad\square \tag{4.3}$$

Let us remark that the dimension of the space (Λ,λ) is the internal multiplicity of the weight $\lambda\in\Pi(\Lambda)$; the PRV theorem states that the external multiplicity of the irreducible compoennt $(\lambda+\Omega)$ of the Kronecker product $(\Lambda)\otimes(\Omega)$ is equal to the dimension of the subspace (Ω,Λ,λ) of (Λ,λ).

B. Klimyk's theorem[10]

For a generic Kronecker product $(\Lambda)\otimes(\Omega)$ in the conditions of the PRV theorem the inequality

$$\dim(\Omega,\Lambda,\lambda)\leqslant\dim(\Lambda,\lambda) \tag{4.4}$$

holds. The following question arises.

What condition ensures that the inequality (4.4) becomes an equality for all the weights $\lambda\in\Pi(\Lambda)$ for which

$$\lambda+\Omega\in\Lambda^+, \tag{4.5}$$

i.e., under what condition does the value of the external multiplicity $\dim(\Omega,\Lambda,\lambda)$ coincide with the internal multiplicity $\dim(\Lambda,\lambda)$?

From the definition of the coefficients $p_i(\lambda)$ it is clear that

$$x_i^{p_i(\lambda)}v\neq 0,\quad x_i^{p_i(\lambda)+1}v=0. \tag{4.6}$$

Thus in order to have

$$\dim(\Omega,\Lambda,\lambda)=\dim(\Lambda,\lambda) \tag{4.7}$$

it is sufficient that for all $i=1,2,\dots,n$, we have

$$p_i(\lambda)\leqslant\langle\Omega,\alpha^\vee{}_i\rangle. \tag{4.8}$$

However, the inequality (4.8) is satisifed (Proposition 3.5) for any $\alpha_i=$ long (short) simple root provided that

$$\langle\Omega,\alpha^\vee{}_i\rangle+1\geqslant\langle\Lambda,\alpha^\vee{}_{hl(hs)}\rangle. \tag{4.9}$$

We can thus state the following proposition.

Proposition 4.1: Let Λ and Ω be two dominant weights of a semisimple Lie algebra L which possess the property

$$\langle\Omega+\rho,\alpha^\vee{}_i\rangle\geqslant\langle\Lambda,\alpha^\vee{}_{hl(hs)}\rangle \tag{4.10}$$

for any long (short) simple root α_i of L. Then the Kronecker product $(\Lambda)\otimes(\Omega)$ admits the decomposition

$$(\Lambda)\otimes(\Omega)=\bigoplus_{\substack{\lambda\in\Pi(\Lambda)\\ \lambda+\Omega\in\Lambda^+}}m_\Lambda(\lambda)(\lambda+\Omega), \tag{4.11}$$

where $m_\Lambda(\lambda)=\dim(\Lambda,\lambda)$ is the internal multiplicity of the weight $\lambda\in\Pi(\Lambda)$. □

Proposition 4.1 is a different formulation of a theorem attributed to Klimyk.[10]

Indeed, Klimyk's theorem[10] states that the equality (4.11) holds provided that

$$\lambda+\Omega+\rho\in\Lambda^+,\ [\text{for any }\Lambda\in\Pi(\Lambda)], \tag{4.12}$$

i.e., provided that $\lambda+\Omega+\rho$ is a dominant weight. Condition (4.12) may also be written as

$$\langle\lambda+\Omega+\rho,\alpha^\vee{}_i\rangle\geqslant 0[\text{for any }\lambda\in\Pi(\Lambda)\text{ and any }i=1,2,\dots,n]. \tag{4.13}$$

However, from condition (4.10) and Proposition 3.1 we obtain

$$\langle \Omega + \rho, \alpha^{\vee}_{i} \rangle \geqslant |\langle \lambda, \alpha^{\vee}_{i} \rangle| \geqslant -\langle \lambda, \alpha^{\vee}_{i} \rangle, \tag{4.14}$$

from which Eq. (4.12) follows.

Let us now prove that, conversely, Eq. (4.10) results from Eq. (4.12), i.e., from the validity of the inequalities

$$\langle \Omega + \rho, \alpha^{\vee}_{i} \rangle \geqslant -\langle \lambda, \alpha^{\vee}_{i} \rangle \tag{4.13'}$$

for any $\lambda \in \Pi(\Lambda)$ and any $i = 1,2,...n$: The validity of Eq. (4.13′) for any $\lambda \in \Pi(\Lambda)$ implies that

$$\langle \Omega + \rho, \alpha^{\vee}_{i} \rangle \geqslant \max_{\lambda \in \Pi(\Lambda)} (-\langle \lambda, \alpha^{\vee}_{i} \rangle). \tag{4.15}$$

Equation (4.15) proves the perfect equivalence between Proposition 3.6 and Klimyk's theorem.[10]

Remark 1: For each particular type of classical Lie algebras, Klimyk has deduced explicit expressions for condition (4.12). Similar expressions for the exceptional Lie algebras have been found by Zaccaria.[23] Equation (4.10) represents a unified form for all these conditions.

The following proposition is important for its applications

Proposition 4.2: If (Λ) is a minuscule representation of a simisimple Lie algebra L, then condition (4.10) is satisfied for any finite-dimensional irreducible representation (Ω) of L.

Proof: Minuscule representations are characterized [Sec. II D, (iv)] by the equation $\langle \Lambda, \alpha^{\vee}_{hs} \rangle = 1$. Recalling Eq. (3.24), the inequality (4.10) becomes

$$\langle \Omega + \rho, \alpha^{\vee}_{i} \rangle = \langle \Omega, \alpha^{\vee}_{i} \rangle + 1 \geqslant 1, \tag{4.16}$$

which is satisfied for any $\lambda \in \Lambda^{+}$. □

Remark 2: Equation (4.10) and its consequence Eq. (4.16) show that the term ρ introduced by Klimyk in Eq. (4.12) is essential for pointing out the special role played by minuscule weights in Kronecker products.

Corollary 4.3: The Kronecker product $(\Lambda) \otimes (\Omega)$, in which (Λ) is a minuscule representation and (Ω) is a finite-dimensional irreducible representation of a semisimple Lie algebra, decomposes into a Clebsch–Gordan series containing the irreducible representations which have as highest weights the set of weights

$$\{\lambda + \Omega | \lambda \in \Pi(\Lambda)\} \cap \Lambda^{+}, \tag{4.17}$$

i.e., the dominant weights that result by translating with Ω the weight diagram of representation (Λ). All the components of the Clebsch–Gordan series have multiplicity 1.

Remark 3: Formula (4.17) for a Kronecker product in which one of the factors is a minuscule representation has also been pointed out by Kass.[24]

Explicit expressions for the sets of weights $\Pi(\Lambda)$ for all semisimple Lie algebras that possess minuscule weights are given in Appendices A and B.

Comment: The following PRV conjecture has recently been proved by Kumar[25]: His proof makes use of the algebraic geometry of Schubert varieties.

The PRV conjectures[25]: Let L be a finite-dimensional complex semisimple Lie algebra with the associated Weyl group W and let (Λ) and (Ω) be two finite-dimensional irreducible representations of L with highest weights Λ and Ω, respectively. Then for any $w \in W$ the irreducible representation $(\overline{\Omega + w\Lambda})$ occurs with a mulitplicity of at least 1 in the product $(\Lambda) \otimes (\Omega)$. We have denoted

$$\overline{\Omega + w\Lambda} = \{\Omega + w\Lambda | w \in W\} \cap \Lambda^{+}. \tag{4.18}$$

In other terms, the PRV conjecture states that

$$N_{CG}(\Lambda, \Omega) \geqslant \#\{\overline{\Omega + w\Lambda} | w \in W\}, \tag{4.19}$$

where $\#\{M\}$ denotes the cardinal number of the set M and $N_{CG}(\Lambda, \Omega)$ the number of terms in the Clebsch–Gordan series of $(\Lambda) \otimes (\Omega)$.

Let us now remark that for Λ_{MW} = minuscule weight Corollary (3.8) of Klimyk's theorem[10] shows that

$$N_{CG}(\Lambda_{MW}, \Omega) = \#\{\overline{\Omega + w\Lambda_{MW}} | w \in W\}, \tag{4.20}$$

i.e., that in this case the inequality of the PRV conjecture becomes an equality.

C. Feingold's theorem

The PRV theorem and the inequalities for the weights of a representation, as derived in Sec. III, can also be used to give a simple proof for the following theorem attributed to Feingold.[12]

Theorem: Let L be a semisimple Lie algebra of rank n. Let Λ, Ω be dominant weights of L $(\Lambda, \Omega \in \Lambda^{+})$ and let (Λ), (Ω) be the corresponding irreducible representations. Let us assume that for the long (short) simple root α_i of L,

$$\langle \Omega, \alpha^{\vee}_{i} \rangle \geqslant \langle \Lambda, \alpha^{\vee}_{hl(hs)} \rangle. \tag{4.21}$$

Then if

$$(\Lambda) \otimes (\Omega) = \bigoplus_{\Gamma \in \Lambda^{+}} m_{\Gamma} \cdot (\Gamma) \tag{4.22}$$

[m_{Γ} = mulitplicity of the representation (Γ)], we also have

$$(\Lambda) \otimes (\Omega + \Lambda_i) = \bigoplus_{\Gamma \in \Lambda^{+}} m_{\Gamma} \cdot (\Gamma + \Lambda_i). \tag{4.23}$$

Proof: The PRV theorem enables us to write the Kronecker product $(\Lambda) \otimes (\Omega)$ using Eqs. (4.3) and (4.2). To derive Eq. (4.24) we have to prove that the Clebsch–Gordan series

$$(\Lambda) \otimes (\Omega + \Lambda_i) = \bigoplus_{\substack{\lambda \in \Pi(\Lambda) \\ \lambda + \Omega + \Lambda_i \in \Lambda^{+}}} (\dim(\Omega + \Lambda_i, \Lambda, \lambda)) \times (\lambda + \Omega + \Lambda_i) \tag{4.24}$$

has the same number of distinct terms as the Clebsch–Gordan series (4.3) and the same multiplicities, i.e., that

$$\{\lambda \in \Pi(\Lambda) | \lambda + \Omega \in \Lambda^{+}\} = \{\lambda \in \Pi(\Lambda) | \lambda + \Omega + \Lambda_i \in \Lambda^{+}\} \tag{4.25}$$

and

$$\dim(\Omega + \Lambda_i, \Lambda, \lambda) = \dim(\Omega, \Lambda, \lambda) \tag{4.26}$$

for any $\lambda \in \Pi(\Lambda)$, for which $\lambda + \Omega \in \Lambda^{+}$ and $\lambda + \Omega + \Lambda_i \in \Lambda^{+}$.

Let us first prove Eq. (4.25); to do that we shall prove that if condition (4.21) is satisfied and if

$$\langle \lambda + \Omega, \alpha^{\vee}_{j} \rangle < 0 \tag{4.27}$$

for any $j = 1,2,...,n$, then

$$\langle \lambda + \Omega + \Lambda_i, \alpha^\vee_j \rangle < 0 \tag{4.28}$$

for any j. For the labels $j \neq i$ the two inequalities are equivalent. Let us assume that $j = i$ and suppose that inequality (4.27) is true and inequality (4.28) is false, i.e., assume that

$$\langle \lambda + \Omega + \Lambda_i, \alpha^\vee_i \rangle = 0. \tag{4.29}$$

Let us observe that from Eqs. (4.29) and (2.11) we obtain

$$\langle \Omega, \alpha^\vee_i \rangle = p_i(\lambda) - q_i(\lambda) - 1 \leqslant p_i(\lambda) \leqslant \langle \Lambda, \alpha^\vee_{hl(hs)} \rangle, \tag{4.30}$$

in contradiction with Eq. (4.22).

Let us now prove that Eq. (4.21) leads to Eq. (4.25), i.e., that for the simple root α_i for which Eq. (4.21) holds,

$$\{v \in (\Lambda, \lambda) | x_i^{\langle \Omega + \Lambda_i + \rho, \alpha^\vee_i \rangle} v = 0\}$$
$$= \{v \in (\Lambda, \lambda) | x_i^{\langle \Omega + \rho, \alpha^\vee_i \rangle} \quad v = 0\}. \tag{4.31}$$

For this simple root α_i,

$$\langle \Omega + \rho, + \Lambda_i, \alpha^\vee_i \rangle > \langle \Omega, \alpha^\vee_i \rangle \geqslant \langle \Lambda, \alpha^\vee_{hl(hs)} \rangle \geqslant p_i(\lambda), \tag{4.32}$$

i.e., for this value of i the two sets in Eq. (4.31) are equal. □

Remark 4: Feingold's theorem[12] has also been obtained by Sakane and Takeuchi (Ref. 17, Theorem 3.4).

Let us assume that in Eqs. (4.21) and (4.22) we take $\Lambda = \Omega$. In this case, Eq. (4.21) can hold only as an equality:

$$\langle \Lambda, \alpha^\vee_i \rangle = \langle \Lambda, \alpha^\vee_{hl(hs)} \rangle, \tag{4.33}$$

with α_i = long (short) simple root. The equality (4.33) can be considered as an equation determining the pairs $\{\alpha_i, \Lambda\}$ which satisfy it.

It is remarkable that the weights Λ that are solutions of Eq. (4.33) are the "Okubo partners"[5,7] of the minuscule weights and —for the algebras B_n and C_n— also the minuscule weights themselves. For the reader's convenience we recall that the "Okubo partner" of a minuscule representation (Λ) is a finite-dimensional representation (Ω) such that the Kronecker product $(\Lambda) \otimes (\Omega)$ decomposes into two irreducible components.

Equation (4.33) has been discussed by Sakane and Takeuchi[17] in connection with the problem of the determination of the Kähler manifolds, which are Hermitian symmetric spaces and are embedded into a projective space by second-degree polynomial defining equations.

In connection with this problem, let us observe that the Okubo partners that resulted above as solutions of Eq. (4.33) have been found by Cavalli *et al.*[15] as solutions of the problem of embedding the Hermitian symmetric spaces as orbits of the highest weight vector of an irreducible representation of a compact semisimple Lie group. Feingold's theorem[12] admits the following immediate generalization.

Proposition 4.4: As before, let Λ and Ω be dominant weights of the semisimple Lie algebra L and let Eq. (4.22) describe the decomposition of the Kronecker product of the representations (Λ) and (Ω) in irreducible components. If for a long (short) root α_i of L the inequality (4.21) is satisfied, then for an $k \in Z^+$, the following decomposition holds:

$$(\Lambda) \otimes (\Omega + k_i) = \bigoplus_{\Gamma \in \Lambda^+} m(\Gamma)(\Gamma + k\Lambda_i). \tag{4.34}$$

Proof: If the dominant weights Λ and Ω satisfy the inequality (4.21), then Λ and $\Omega + k\Lambda_i$ $(k \in Z^+)$ will satisfy it as well: The proof of Eq. (4.34) is identical to the proof of Eq. (4.23).

Assume now that Λ is a minuscule weight, i.e., that $\langle \Lambda, \alpha^\vee_{hl} \rangle = \langle \Lambda, \alpha^\vee_{hs} \rangle = 1$. We obtain the following corollary to Feingold's theorem.[12]

Corollary 4.5: Let L be a semisimple Lie algebra of rank n and let Λ be a minuscule weight of L. Let $\Omega \in \Lambda^+$ be a dominant weight of L subject to the condition $\langle \Omega, \alpha^\vee_i \rangle \geqslant 1$, but otherwise arbitrary. Then if the Kronecker product $(\Lambda) \otimes (\Omega)$ admits the decomposition (4.22), the Kronecker product $(\Lambda) \otimes (\Omega + k\Lambda_i)$ admits the decomposition (4.34) for any $i = 1,2,...n$ and for any $k \in Z^+$. In particular, assuming that Ω is the fundamental weight Λ_i and Λ is a minuscule weight, condition (4.21) is satisfied, so that the decomposition

$$(\Lambda) \otimes (\Lambda_i) = \bigoplus_{\Gamma \in \Lambda^+} m_\Gamma(\Gamma) \tag{4.35}$$

implies the validity, for any $i = 1,2,...,n$ of the decomposition

$$(\Lambda) \otimes ((k+1)\Lambda_i) = \bigoplus_{\Gamma \in \Lambda^+} m(\Gamma + k\Lambda_i). \tag{4.36}$$

D. Dynkin's theorem[11]

In this section we point out that the PRV theorem can be used to prove the following result attributed to Dynkin,[11] which gives the expression of (the highest weight of) a representation—different from the Cartan (stretched) representation—and which is contained in any product of two irreducible representations (Λ) and (Ω) of a semisimple Lie algebra.

The importance of Dynkin's result is that the Cartan and Dynkin representations are the *only* irreducible representations that appear in the Kronecker product of *any* nontrivial two irreducible representations of *any* semisimple Lie algebra. In other words, if the Clebsch–Gordan series of a Kronecker product $(\Lambda) \otimes (\Omega)$ has length 2, then this series is composed of the Cartan and Dynkin representations only.

Let us state (a part of) Dynkin's theorem.

Theorem[11]: Let (Λ) and (Ω) be two finite-dimensional irreducible representations of a semisimple Lie algebra L, which are labeled by the dominant weights Λ and Ω, respectively. Let $\alpha_1, \alpha_2, ..., \alpha_k$ be a system of simple roots of L which "connects" the highest weights Λ and Ω, i.e., which possesses the properties

$$\langle \Lambda, \alpha^\vee_1 \rangle \neq 0, \quad \langle \Lambda, \alpha^\vee_2 \rangle = \cdots = \langle \Lambda, \alpha^\vee_k \rangle = 0, \tag{4.37}$$

$$\langle \alpha_i, \alpha^\vee_{i+1} \rangle \neq 0, \quad (i = 1,2,...,k-1), \tag{4.38}$$

$$\langle \alpha_i, \alpha^\vee_j \rangle = 0, \quad \text{if } j > i+1 \text{ (for any } i), \tag{4.39}$$

$$\langle \Omega, \alpha^\vee_1 \rangle = \langle \Omega, \alpha^\vee_2 \rangle = \cdots = \langle \Omega, \alpha^\vee_{k-1} \rangle = 0, \tag{4.40}$$

$$\langle \Omega, \alpha^\vee_k \rangle \neq 0. \tag{4.41}$$

Then the Kronecker product $(\Lambda) \otimes (\Omega)$ contains the representation $(\Omega + \Lambda')$, where Λ' is the weight defined by

$$\Lambda' = \Lambda - \alpha_1 - \alpha_2 - \cdots - \alpha_k. \tag{4.42}$$

The multiplicity of the representation $(\Omega + \Lambda')$ in the decomposition of $(\Lambda) \otimes (\Omega)$ is 1.

We shall call $\Omega + \Lambda'$ a Dynkin weight and $(\Omega + \Lambda')$ a Dynkin representation.

The *proof* of the theorem will use the following lemma, attributed to Dynkin and the PRV theorem.

Lemma[11]: Let (Λ) be a finite-dimensional irreducible representation with highest weight Λ, v_Λ the corresponding highest weight vector, and $\alpha_1,\alpha_2,...,\alpha_k$ a chain of simple roots with the properties (4.37)–(4.39). Then using the notations (4.1), we have

$$y_{\alpha_k} y_{\alpha_{k-1}} \cdots y_{\alpha_2} y_{\alpha_1} v_\Lambda \neq 0 \tag{4.43}$$

and

$$y_{\alpha_{i_k}} y_{\alpha_{i_{k-1}}} \cdots y_{\alpha_{i_2}} y_{\alpha_{i_1}} v_\Lambda = 0, \tag{4.44}$$

if the permutation $(i_1,i_2,...,i_k) \neq (1,2,...,k)$. □

It is easy to see that the weight $\Lambda' \in \Pi(\Lambda)$ defined by Eq. (4.42) has internal multiplicity equal to 1. The weight (4.42) results by applying a product of Weyl reflections (which conserves the internal multiplicity) to the highest weight.

Let us also observe that the coefficients $p_j(\Lambda')$ associated [cf. Eq. (2.10)] to the weight Λ' defined by Eq. (4.42) have the values

$$p_1(\Lambda') = p_2(\Lambda') = \cdots p_{k-1}(\Lambda') = 0, \tag{4.45}$$

$$p_k(\Lambda') = 1, \tag{4.46}$$

$$p_{k+1}(\Lambda') = p_{k+2}(\Lambda') = \cdots = p_n(\Lambda') = 0. \tag{4.47}$$

Using the commutation relations (4.1) we obtain

$$x_{\alpha_k}(y_{\alpha_k} y_{\alpha_{k-1}} \cdots y_{\alpha_1} v_\Lambda) = -\langle \alpha_{k-1}, \alpha^\vee{}_k \rangle y_{\alpha_{k-1}} y_{\alpha_{k-2}} \cdots y_{\alpha_1} v_\Lambda \neq 0 \tag{4.48}$$

because of Eq. (4.39). Again using Eqs. (4.1) we obtain

$$x^2{}_{\alpha_k}(y_{\alpha_k} y_{\alpha_{k-1}} \cdots y_{\alpha_1} v_\Lambda) = 0. \tag{4.49}$$

This proves Eq. (4.46). Equations (4.45) and (4.47) result, in a similar way, from using Eqs. (4.1), (4.39), and (4.44).

Let us now consider the weight Λ' defined by Eqs. (4.37)–(4.42); let (Λ,Λ') be its weight vectors subspace (cf. Sec. IV A). We have to prove that for any $v \in (\Lambda,\Lambda')$,

$$x_j^{\langle \Omega + \rho, \alpha^\vee{}_j \rangle} v = 0, \quad \text{for any } j = 1,2,...,n. \tag{4.50}$$

For $j < k$ Eq. (4.41) leads to

$$\langle \Omega + \rho, \alpha^\vee{}_j \rangle = 1 \tag{4.51}$$

and the equalities (4.45) imply that $x_j v = 0$ for $j = 1,2,...,k-1$. Let us consider $j = k$ and observe that because $\langle \Omega, \alpha^\vee{}_k \rangle \neq 0$, we obtain

$$\langle \Omega + \rho, \alpha^\vee{}_k \rangle \geqslant 2 \tag{4.52}$$

and, since $p_k(\Lambda') = 1$, this implies

$$x_k^{\langle \Omega + \rho, \alpha^\vee{}_k \rangle} v = 0. \tag{4.53}$$

Finally, for $j > k$, Eqs. (4.47) hold and since in this case $\langle \Omega + \rho, \alpha^\vee{}_j \rangle \geqslant 1$, Eq. (4.50) is again satisfied and the proof is complete. □

V. KRONECKER PRODUCTS WITH MINUSCULE REPRESENTATIONS

The minimum length of the Clebsch–Gordan series of the Kronecker product of two nondegenerate finite-dimensional irreducible representations of a semisimple Lie algebra is 2. As already stated, the Clebsch–Gordan series of length 2 contain only the two "standard" representations : the Cartan and the Dynkin representations.

It is also to be expected that the Kronecker factors leading to Clebsch–Gordan decompositions of length 2 be representations with remarkable properties.

In the present section we prove that for a classical Lie algebra, the number $N_{\rm CG}(\Lambda,\Omega)$ of terms in the Clebsch–Gordan series of a Kronecker product $(\Lambda \otimes (\Omega)$ attains the minimum $N_{\rm CG}(\Lambda,\Omega) = 2$ if Λ is a minuscule weight and Ω is an integer multiple of a fundamental weight associated with the minuscule weight Λ and called its "Okubo partner."

Proposition 5.1: (i) If a Kronecker product $(\Lambda) \otimes (\Omega)$ of two finite-dimensional irreducible representations of a classical semisimple Lie algebra decomposes into two irreducible components, then one of the factors is a minuscule representation.

(ii) For the algebras B_n, C_n, and D_n the Okubo partners (Ω) associated with the minuscule representations (Λ) are shown in Table II.

Similar results are valid for the algebras of type A_n; they result from the direct application of the multiplication of Young tableaux and are pointed out in Table I.

The proof of Proposition 5.1 is based on the following lemmas.

Lemma 5.2: Let Λ, Ω, Ω' be dominant weights. The decomposition of the Kronecker product $(\Lambda) \otimes (\Omega + \Omega')$ contains any representation $(\lambda + \Omega + \Omega')$ with $\lambda \in \Pi(\Lambda)$, $\lambda + \Omega + \Omega' \in \Lambda^+$, with a multiplicity larger or equal than the multiplicity of the representation $(\lambda + \Omega) \in \Lambda^+$ in the decomposition of $(\Lambda) \otimes (\Omega)$.

Proof: Along with the notations of Sec. IV A, let $v \in (\Lambda,\lambda)$ and $\Omega \in \Lambda^+$ such that

$$x_j^{\langle \Omega + \rho, \alpha^\vee{}_j \rangle} v = 0, \quad \text{for any } j = 1,2,...,n. \tag{5.1}$$

Equation (5.1) implies that for any $\Omega' \in \Lambda^+$,

$$x_j^{\langle \Omega + \Omega' + \rho, \alpha^\vee{}_j \rangle} v = 0, \quad \text{for any } j = 1,2,..,n, \tag{5.2}$$

from which, from the PRV theorem, the stated inequality between the multiplicities results. □

Corollary 5.3: Let $N_{\rm CG}(\Lambda,\Omega)$ be the number of terms in the Clebsch–Gordan decomposition of the Kronecker product $(\Lambda) \otimes (\Omega)$ and let Λ_i be fundamental weights of the semisimple Lie algebra L. Then

$$N_{\rm CG}(\Lambda_i,\Lambda_j) \leqslant N_{\rm CG}(\Lambda_i + \Lambda_k,\Lambda_j). \tag{5.3}$$

TABLE II. The Okubo partners (Ω) associated with the minuscule representation (Λ) for the algebras B_n, C_n, and D_n.

Algebra	Minuscule weights	Okubo partners
B_n	Λ_n	$k\Lambda_1 (k = 1,2,...)$
C_n	Λ_1	$k\Lambda_n (k = 1,2,...)$
D_n	Λ_1	$k\Lambda_{n-1}, k\Lambda_n (k = 1,2,...)$
	Λ_{n-1}, Λ_n	$k\Lambda_1 (k = 1,2,...)$

The Kronecker products with the shortest Clebsch–Gordan series therefore have to be sought among the products of fundamental representations.

Lemma 5.4: Let (Λ_i) and (Λ_j) with $1 \leqslant i < j < n$ be two fundamental representations of a Lie algebra of type C_n. The decomposition of the Kronecker product $(\Lambda_i) \otimes (\Lambda_j)$ contains the representations (Λ_{j-i}), which is a Dynkin representation only for $i = 1$ and $j = n$.

Proof: Inspection of the Cartan matrix of C_n points out that (denoting again simple roots by α_i)

$$\alpha_1 + \alpha_2 + \cdots + \alpha_k = \Lambda_1 + \Lambda_k - \Lambda_{k+1} . \quad (5.4)$$

Equation (5.4) proves the existence of the following chain, with $k < n$ and a decreasing sequence of the labels of the simple roots:

$$\Lambda_k \xrightarrow{\alpha_k} \Lambda_{k-1} - \Lambda_k + \Lambda_{k+1} \xrightarrow{\alpha_{k-1}} \Lambda_{k-2} - \Lambda_{k-1} + \Lambda_{k+1} \xrightarrow{\alpha_{k-2}} \cdots \xrightarrow{\alpha_1} -\Lambda_1 + \Lambda_{k-1}. \quad (5.5)$$

Using Dynkin's lemma (Sec. IV D) we can prove that each element of this chain is either the start or end of an α_i string, i.e., no element is "traversed" by any other string. This is essential for the application of the PRV theorem. We thus have

$$\Lambda_k - (\alpha_k + \alpha_{k-1} + \cdots + \alpha_2 + \alpha_1) = -\Lambda_1 + \Lambda_{k+1} . \quad (5.6)$$

Similarly,

$$-\Lambda_1 + \Lambda_{k+1} - (\alpha_{k+1} + \alpha_k + \cdots + \alpha_2) = -\Lambda_2 + \Lambda_{k+2} , \quad (5.7)$$

$$-\Lambda_2 + \Lambda_{k+2} - (\alpha_{k+2} + \alpha_{k+1} + \cdots + \alpha_3) = -\Lambda_3 + \Lambda_{k+3} , \quad (5.8)$$

and, finally, from

$$\alpha_{n-k} + \alpha_{n-k+1} + \cdots + \alpha_{n-1} = -\Lambda_n + \Lambda_{n-1} + \Lambda_{n-k} - \Lambda_{n-k-1} \quad (5.9)$$

we obtain

$$-\Lambda_{n-k-1} + \Lambda_{n-1} - (\alpha_{n-1} + \alpha_{n-2} + \cdots + \alpha_{n-k}) = -\Lambda_{n-k} + \Lambda_n . \quad (5.10)$$

The following weights [which belong to $\Pi(\Lambda_k)$]:

$$\Lambda_k, \Lambda_{k+1} - \Lambda_1, \quad \Lambda_{k+2} - \Lambda_2, \ldots, \Lambda_n - \Lambda_{n-k} \quad (5.11)$$

are, as a result of the same Dynkin lemma, either the start or the end of an α_i string. The representations Λ_k of C_n being self-contragredient, the same property is true for the following set of weights which belong to $\Pi(\Lambda_k)$:

$$-\Lambda_k, \quad -\Lambda_{k+1} + \Lambda_1, \quad -\Lambda_{k+2} + \Lambda_2, \ldots, -\Lambda_n + \Lambda_{n-k} . \quad (5.12)$$

Therefore, application of the PRV theorem gives

$$(\Lambda_k) \otimes (\Lambda_k) \ni (\Lambda_0) ,$$
$$(\Lambda_k) \otimes (\Lambda_{k+1}) \ni (\Lambda_1) ,$$
$$\cdots$$
$$(\Lambda_k) \otimes (\Lambda_n) \ni (\Lambda_{n-k}) , \quad (5.13)$$

where $(A) \otimes (B) \ni (C)$ means that representation (C) belongs to the Clebsch–Gordan series of the Kronecker product $(A) \otimes (B)$.

By applying this procedure successively to the representations $(\Lambda_1), (\Lambda_2), \ldots, (\Lambda_n)$ we obtain the general result $(k + i \leqslant n)$:

$$(\Lambda_k) \otimes (\Lambda_{k+i}) \ni (\Lambda_i) . \quad (5.14)$$

It is easy to prove that Λ_i is not a Dynkin weight if $k \neq 1$ and $k + i \neq n$. We have, indeed,

$$\Lambda_k + \Lambda_{k+i} - (\alpha_k + \alpha_{k+1} + \cdots + \alpha_{k+i}) = \Lambda_{k-1} + \Lambda_{k+i+1} \neq \Lambda_i . \quad (5.15)$$

If $k = 1$ and $k + i = n$, then $\alpha_1 + \alpha_2 + \cdots + \alpha_n = \Lambda_1 - \Lambda_{n-1} + \Lambda_n$ and $\Lambda_1 + \Lambda_n - \Sigma_{i=1}^{n} \alpha_i = \Lambda_{n-1}$ is a Dynkin weight.

Lemma 5.5: Let (Λ_i) and (Λ_j) with $1 \leqslant i < j < n - 1$ be fundamental representations of a Lie algebra of type B_n. The decomposition of the product $(\Lambda_i) \otimes (\Lambda_j)$ contains the representation (Λ_{j-i}), which is not a Dynkin representation. The products $(\Lambda_n) \otimes (\Lambda_i)$ $(i = 1,2,\ldots,n-1)$ contain the representation (Λ_n), which is not a Dynkin representation except for $i = 1$. We also have $(\Lambda_n) \otimes (\Lambda_n) \ni (\Lambda_0)$.

Proof: Inspection of the Cartan matrix of B_n and a reasoning similar to Lemma 5.4 show that the weights

$$-\Lambda_k, \quad -\Lambda_{k+1} + \Lambda_1, \quad -\Lambda_{k+2} + \Lambda_2, \ldots, -\Lambda_{n-1} + \Lambda_{n-k-1} \quad (5.16)$$

are either the end or beginning of α_i strings and are not traversed by any string. Applying the PRV theorem to these weights proves that

$$(\Lambda_k) \otimes (\Lambda_{k+i}) \ni (\Lambda_i) , \quad \text{for } i = 0,1,2,\ldots,n-k-1 . \quad (5.17)$$

It can be proved that none of the weights $(\Lambda_0), (\Lambda_1), \ldots, (\Lambda_{n-k-1})$ are Dynkin weights for the corresponding Kronecker products.

Let us now recall that Λ_n is a minuscule weight for B_n. All the weights of representation (Λ_n) have multiplicity 1 and each weight is either the beginning or end of α_i strings. Observing that

$$\Lambda_n - (\alpha_n + \alpha_{n-1} + \cdots + \alpha_1) = \Lambda_n - \Lambda_1 ,$$
$$\Lambda_n - \Lambda_1 - (\alpha_n + \alpha_{n-1} + \cdots + \alpha_2) = \Lambda_n - \Lambda_2 ,$$
$$\cdots$$
$$\Lambda_n - \Lambda_{n-2} - (\alpha_n + \alpha_{n-1}) = \Lambda_n - \Lambda_{n-1} , \quad (5.18)$$

and applying the PRV theorem we obtain

$$(\Lambda_n) \otimes (\Lambda_1) \ni (\Lambda_n) = \text{Dynkin representation for } (\Lambda_n) \otimes (\Lambda_1) ,$$
$$(\Lambda_n) \otimes (\Lambda_i) \ni (\Lambda_n) \neq \text{Dynkin representation for } (\Lambda_n) \otimes (\Lambda_i) , \quad \text{for } i = 2,3,\ldots,n-1.$$

Since $-\Lambda_n \in \Pi(\Lambda_n)$ we have $(\Lambda_n) \otimes (\Lambda_n) \ni (\Lambda_0)$. □

Lemma 5.6: Let (Λ_i) and (Λ_j) with $1 \leqslant i < j < n - 2$ be two fundamental representations of a Lie algebra of type D_n $(n \geqslant 4)$. The Clebsch–Gordan series of the Kronecker product $(\Lambda_i) \otimes (\Lambda_j)$ contains the representation (Λ_{j-i}), which is not a Dynkin representation. We also have, for $i\leqslant n-2$,

$$(\Lambda_n)\otimes(\Lambda_i)\ni(\Lambda_n)\,,\quad \text{if } i=2k\,, \tag{5.19}$$

$$(\Lambda_n)\otimes(\Lambda_i)\ni(\Lambda_{n-1})\,,\quad \text{if } i=2k+1\,, \tag{5.20}$$

$$(\Lambda_{n-1})\otimes(\Lambda_i)\ni(\Lambda_{n-1})\,,\quad \text{if } i=2k\,, \tag{5.21}$$

$$(\Lambda_{n-1})\otimes(\Lambda_i)\ni(\Lambda_n)\,,\quad \text{if } i=2k+1\,, \tag{5.22}$$

and

$$(\Lambda_{n-1})\otimes(\Lambda_n)\ni(\Lambda_0)\,. \tag{5.23}$$

Proof: In a similar way as for the algebras C_n and B_n we can prove that for $i=0,1,2,\ldots,n-2$,

$$(\Lambda_k)\otimes(\Lambda_{k+i})\ni(\Lambda_i)\,. \tag{5.24}$$

Let us consider now the spinorial representations (Λ_{n-1}) and (Λ_n). The following equalities hold:

$$\Lambda_n-\Big(\sum_{i=1}^{n}\alpha_i\Big)=\Lambda_{n-1}-\Lambda_1\,,$$
$$\Lambda_{n-1}-\Big(\sum_{i=1}^{n-1}\alpha_i\Big)=\Lambda_n-\Lambda_1\,, \tag{5.25}$$

$$\Lambda_{n-1}-\Lambda_1-\Big(\sum_{i=2}^{n-1}\alpha_i\Big)=\Lambda_n-\Lambda_2\,, \tag{5.26}$$

$$\Lambda_n-\Lambda_1-\Big(\sum_{i=2}^{n-2}\alpha_i+\alpha_n\Big)=\Lambda_{n-1}-\Lambda_2\,, \tag{5.27}$$

$$\Lambda_n-\Lambda_2-\Big(\sum_{i=3}^{n-2}\alpha_i+\alpha_n\Big)=\Lambda_{n-1}-\Lambda_3\,, \tag{5.28}$$

$$\Lambda_{n-1}-\Lambda_2-\Big(\sum_{i=3}^{n-1}\alpha_i\Big)=\Lambda_n-\Lambda_3\,, \tag{5.29}$$

$$\cdots\cdots.$$

Application of the PRV theorem gives

$$(\Lambda_n)\otimes(\Lambda_1)\ni(\Lambda_{n-1})\,,\quad(\Lambda_{n-1})\otimes(\Lambda_1)\ni(\Lambda_n)\,. \tag{5.30}$$

In Eqs. (5.30) (Λ_{n-1}) and (Λ_n) are Dynkin representations in the products $(\Lambda_n)\otimes(\Lambda_1)$ and $(\Lambda_{n-1})\otimes(\Lambda_1)$, respectively. Relations (5.19)–(5.22) result from Eqs. (5.26) – (5.29) by application of the PRV theorem. The representations (Λ_{n-1}) and (Λ_n) are not Dynkin representations in the corresponding Kronecker products.

To prove Proposition 5.1 let us recall that the decomposition of the Kronercker product of any two irreducible finite-dimensional representations contains two "standard" irreducible components : the Cartan (stretched) and Dynkin representations. The Kronecker products whose decompositions contain a representation that is neither Cartan nor Dynkin have to be excluded in our search for pairs of fundamental weights $\{\Lambda_i,\Lambda_j\}$, for which the product $(\Lambda_i)\otimes(\Lambda_j)$ decomposes into two irreducible terms. From the previous analysis these pairs are the following:

$\{\Lambda_1,\Lambda_n\}$, for B_n, C_n;

$\{\Lambda_1,\Lambda_{n-1}\}$, $\{\Lambda_1,\Lambda_n\}$, for D_n.

Let us now recall that Λ_n is a minuscule weight for B_n; Λ_1 is a minuscule weight for C_n; and Λ_1, Λ_{n-1}, and Λ_n are minuscule weights for D_n. Using Corollary 4.5 to Feingold's theorem (Sec. IV C) this enables us to conclude that

$$N_{CG}(\Lambda_1,\Lambda_n)=N_{CG}(k\Lambda_1,\Lambda_n)=2,$$
$$(k=1,2,\ldots)\ \text{for}\ B_n\,,$$
$$N_{CG}(\Lambda_n,\Lambda_1)=N_{CG}(k\Lambda_n,\Lambda_1)=2,$$
$$(k=1,2,\ldots)\ \text{for}\ C_n\,,$$
$$N_{CG}(\Lambda_1,\Lambda_{n-1})=N_{CG}(k\Lambda_1,\Lambda_{n-1})$$
$$=N_{CG}(\Lambda_1,k\Lambda_{n-1})=2\,,(k=1,2,\ldots)\,,$$
$$N_{CG}(\Lambda_1,\Lambda_n)=N_{CG}(k\Lambda_1,\Lambda_n)=N_{CG}(\Lambda_1,k\Lambda_n)=2,$$
$$(k=1,2,\ldots)\ \text{for}\ D_n$$

□

APPENDIX A: THE WEIGHTS OF THE MINUSCULE REPRESENTATIONS OF THE CLASSICAL LIE ALGEBRAS

In this section we give analytic expressions for the weights of the minuscule representations of the classical Lie algebras of arbitrary rank.

1. Algebras of type A_n

For the algebras of type A_n the minuscule weights are all the fundamental weights $\Lambda_1,\Lambda_2,\ldots,\Lambda_n$. Using Dynkin's ladder procedure for the derivation of the weights of a representation from its highest weight and Eq. (2.2) we obtain by induction the analytic expression for the weights of the fundamental representations (Λ_k) $(k=1,2,\ldots,n)$ of the Lie algebra A_n; these weights are

$$-\Lambda_{i_1}+\Lambda_{i_1+1}-\Lambda_{i_2}+\Lambda_{i_2+1}-\cdots-\Lambda_{i_k}+\Lambda_{i_k+1}\,, \tag{A1}$$

with $0\leqslant i_1<i_2<\cdots<i_k\leqslant n$ and the conventions $\Lambda_0=0$ and $\Lambda_{n+1}=0$.

In orthogonal coordinates $\epsilon_1,\epsilon_2,\ldots,\epsilon_n,\epsilon_{n+1}$ the expressions of the weights of the minuscule representation (Λ_k) of A_n are

$$\epsilon_{i_1}+\epsilon_{i_2}+\cdots+\epsilon_{i_k}-\frac{k}{n+1}\sum_{j=1}^{n+1}\epsilon_j\,, \tag{A2}$$

with $1\leqslant i_2,\cdots<i_k\leqslant n+1$. The number of weights of representation (Λ_k) of the algebra A_n is thus $\binom{n+1}{k}$, a well-known result.

Let us also observe that the weights of the representation (Λ_{n+1-k}) are obtained from those of the conjugated representation (Λ_k) by reversing the signs of the weights of the last representation.

2. Algebras of type B_n

Algebras of type B_n admit only one minuscule weight: Λ_n. Again using the step-down procedure, we obtain the following set of weights for the representation (Λ_n):

$$\pm[\Lambda_{i_1}-\Lambda_{i_2}+\Lambda_{i_3}-\Lambda_{i_4}+\cdots$$
$$+(-1)^{k-1}\Lambda_{i_k}+(-1)^k\Lambda_n]\,, \tag{A3}$$

with $1\leqslant i_1<i_2<\cdots<i_k\leqslant n-1$ and $k=0,1,2,\ldots,n-1$, where, for $k=0$, the expression of the weights between the square brackets reduces to zero.

The expression of the set of weights (A3) in orthogonal coordinates ϵ_i $(i=1,2,\ldots,n)$ is

$$\tfrac{1}{2}[\pm\epsilon_1\pm\epsilon_2\pm\epsilon_3\pm\cdots\pm\epsilon_n], \tag{A4}$$

without any coherence between the plus and minus signs.

The number of weights of representation (Λ_n), calculated from Eq. (A3), is

$$2\left[\binom{n-1}{0}+\binom{n-1}{1}+\cdots+\binom{n-1}{n-1}\right]=2^n.$$

3. Algebras of type C_n

Algebras of type C_n have only one minuscule weight: Λ_1. The set of weights of representation (Λ_1) is

$$\{\pm[-\Lambda_i+\Lambda_{i+1}]\,|\,i=0,1,2,\dots,n-1\}, \tag{A5}$$

with the convention $\Lambda_0=0$.

In orthogonal coordinates the set of weights (A5) has the expression

$$\{\pm\epsilon_i\,|\,i=1,2,\dots,n\}. \tag{A6}$$

4. Algebras of type D_n

Algebras of type D_n have three minuscule weights: Λ_1, Λ_{n-1}, and Λ_n. The set of weights of the representation (Λ_1) is

$$\{\pm[\Lambda_i-\Lambda_{i-1}],\quad i=1,2,\dots,n-2,n,$$
$$\pm[\Lambda_n+\Lambda_{n-1}-\Lambda_{n-2}]\}, \tag{A7}$$

with $\Lambda_0=0$; altogether, $2l$ weights. In terms of orthogonal coordinates, the set of weights (A7) is

$$\{\pm\epsilon_i\,|\,i=1,2,\dots,n\}. \tag{A8}$$

The set of weights of the representation (Λ_n) for n odd is

$$\left\{(-1)^{i_1+i_2+\cdots+i_k+n-k}\left[\sum_{j=1}^{k}(-1)^j\Lambda_{i_j}+(-1)^{k+1}\Lambda_n\right]\,\middle|\,\begin{matrix}1\leqslant i_1<i_2<\cdots<i_k\leqslant n-2,\\ k=0,1,2,\dots,n-2\end{matrix}\right\},$$
$$\cup\left\{(-1)^{i_1+i_2+\cdots+n+k+1}\left[\sum_{j=1}^{k}(-1)^j\Lambda_{i_j}+(-1)^{k+1}\Lambda_{n-1}\right]\,\middle|\,\begin{matrix}1\leqslant i_1<i_2<\cdots<i_k\leqslant n-2,\\ k=0,1,2,\dots,n-2\end{matrix}\right\}. \tag{A9}$$

In terms of orthogonal coordinates, the set of weights (A9) becomes

$$\frac{1}{2}\sum_{i=1}^{n}\epsilon_i,\quad \frac{1}{2}\sum_{i=1}^{n}(-1)^{\delta_{is}}\epsilon_i,\quad (s=1,2,\dots,n),\quad \frac{1}{2}\sum_{i=1}^{n}(-1)^{\delta_{is}+\delta_{it}}\epsilon_i,\quad (s,t=1,2,\dots,n;\ s<t). \tag{A10}$$

The set of weights of the representation (Λ_{n-1}) for n odd is obtained from the set of weights (A9) by reversing, for the same n, the sign of each weight.

The set of weights of representation (Λ_n) for n even is

$$\left\{\pm\left[\sum_{j=1}^{k}(-1)^j\Lambda_{i_j}+(-1)^{k+1}\Lambda_n\right]\,\middle|\,\begin{matrix}1\leqslant i_1<i_2<\cdots<i_k\leqslant n-2,\\ i_1+i_2+\cdots+i_k+n\equiv 0,\quad (\mathrm{mod}\ 2),\\ k=0,1,2,\dots,n-2\end{matrix}\right\},$$
$$\cup\left\{\pm\left[\sum_{j=1}^{k}(-1)^j\Lambda_{i_j}+(-1)^{k+1}\Lambda_{n-1}\right]\,\middle|\,\begin{matrix}1\leqslant i_1<i_2<\cdots<i_k\leqslant n-2\\ i_1+i_2+\cdots+i_k+n-1\equiv 0,\quad (\mathrm{mod}\ 2)\\ k=0,1,2,\dots,n-2\end{matrix}\right\}. \tag{A11}$$

The set of weights of the representation (Λ_{n-1}) for n even is obtained from the set of weights (A11) by permutating Λ_n with Λ_{n-1} and keeping the rest of the terms and conditions unchanged.

The sets (A10) and (A11) contain

$$2\left[\binom{n-2}{0}+\binom{n-2}{1}+\cdots+\binom{n-2}{n-2}\right]=2^{n-1}$$

terms.

APPENDIX B: THE WEIGHTS OF THE MINUSCULE REPRESENTATIONS OF THE EXCEPTIONAL LIE ALGEBRAS

The only exceptional semisimple Lie algebras that admit minuscule representations are E_6 and E_7. Using Bourbaki's notations, the algebra E_6 admits the minuscule representations (Λ_1) and (Λ_6); the algebra E_7 admits the minuscule representation (Λ_7).

We shall enumerate the weights belonging to a given representation by separating with solidi the weights belonging to different levels of the representation; the various levels of a given representation are presented in sequential order, beginning with the level of the highest weight; and different weights within the same level are separated by commas.

For the algebra E_6 we shall present only the list of weights of representation (Λ_1) because the list of weights of (Λ_6) is obtained from the list of weights of (Λ_1) by simply changing the signs of the former.

The weights of representation (Λ_1) are listed below. Recall that $\dim(\Lambda_1)=\dim(\Lambda_6)=27$:

$$\Lambda_1/-\Lambda_1+\Lambda_3/-\Lambda_3+\Lambda_4/\Lambda_2-\Lambda_4+\Lambda_5/-\Lambda_2+\Lambda_5,\Lambda_2-\Lambda_5+\Lambda_6/-\Lambda_2+\Lambda_4-\Lambda_5+\Lambda_6,$$
$$\Lambda_2-\Lambda_6/\Lambda_3-\Lambda_4+\Lambda_6,\quad -\Lambda_2+\Lambda_4-\Lambda_6/\Lambda_1-\Lambda_3+\Lambda_6,\quad \Lambda_3-\Lambda_4+\Lambda_5-\Lambda_6/-\Lambda_1+\Lambda_6,$$

$\Lambda_1-\Lambda_3+\Lambda_5-\Lambda_6,\quad \Lambda_3-\Lambda_5/-\Lambda_1+\Lambda_5-\Lambda_6,\quad \Lambda_1-\Lambda_3+\Lambda_4-\Lambda_5/-\Lambda_1+\Lambda_4-\Lambda_5,$

$\Lambda_1+\Lambda_2-\Lambda_4/-\Lambda_1+\Lambda_2+\Lambda_3-\Lambda_4,\quad \Lambda_1-\Lambda_2/\Lambda_2-\Lambda_3,$

$-\Lambda_1-\Lambda_2+\Lambda_3/-\Lambda_2-\Lambda_3+\Lambda_4/-\Lambda_4+\Lambda_5/-\Lambda_5+\Lambda_6/-\Lambda_6/.$

This table of weights and the corresponding table of weights of representation (Λ_6) (obtained from the previous by reversing signs) allow us, using Corollary 4.3, to write the following Kronecker products for the Lie algebra E_6:

$$(\Lambda_1)\otimes(\Lambda_1)=(2\Lambda_1)\oplus(\Lambda_3)\oplus(\Lambda_6),\quad (\Lambda_1)\otimes(\Lambda_2)=(\Lambda_1+\Lambda_2)\oplus(\Lambda_5)\oplus(\Lambda_1),$$
$$(\Lambda_1)\otimes(\Lambda_3)=(\Lambda_1+\Lambda_3)\oplus(\Lambda_4)\oplus(\Lambda_1+\Lambda_6)\oplus(\Lambda_2),$$
$$(\Lambda_1)\otimes(\Lambda_4)=(\Lambda_1+\Lambda_4)\oplus(\Lambda_2+\Lambda_5)\oplus(\Lambda_3+\Lambda_6)\oplus(\Lambda_1+\Lambda_2)\oplus(\Lambda_5),$$
$$(\Lambda_1)\otimes(\Lambda_5)=(\Lambda_1+\Lambda_5)\oplus(\Lambda_2+\Lambda_6)\oplus(\Lambda_3)\oplus(\Lambda_6),\quad (\Lambda_1)\otimes(\Lambda_6)=(\Lambda_1+\Lambda_6)\oplus(\Lambda_2)\oplus(\Lambda_0),$$
$$(\Lambda_6)\otimes(\Lambda_2)=(\Lambda_2+\Lambda_6)\oplus(\Lambda_3)\oplus(\Lambda_6),\quad (\Lambda_6)\otimes(\Lambda_3)=(\Lambda_3+\Lambda_6)\oplus(\Lambda_1+\Lambda_2)\oplus(\Lambda_5)\oplus(\Lambda_1),$$
$$(\Lambda_6)\otimes(\Lambda_4)=(\Lambda_4+\Lambda_6)\oplus(\Lambda_2+\Lambda_3)\oplus(\Lambda_1+\Lambda_5)\oplus(\Lambda_2+\Lambda_6)\oplus(\Lambda_3),$$
$$(\Lambda_6)\otimes(\Lambda_5)=(\Lambda_5+\Lambda_6)\oplus(\Lambda_4)\oplus(\Lambda_1+\Lambda_6)\oplus(\Lambda_2),\quad (\Lambda_6)\otimes(\Lambda_6)=(2\Lambda_6)\oplus(\Lambda_5)\oplus(\Lambda_1),$$

as well as, similarly, any other Kronecker product of the form $(\Lambda_1)\otimes(\Omega)$ or $(\Lambda_6)\otimes(\Omega)$ for any dominant weight Ω of E_6.

For the algebra E_7 the minuscule representation (Λ_7) of the Lie algebra E_7 ($\dim(\Lambda_7)=56$) has the following set of weights:

$\Lambda_7/\Lambda_6-\Lambda_7/\Lambda_5-\Lambda_6/\Lambda_4-\Lambda_5/\Lambda_2+\Lambda_3-\Lambda_4/-\Lambda_2+\Lambda_3,\quad \Lambda_1+\Lambda_2-\Lambda_3/\Lambda_1-\Lambda_2-\Lambda_3+\Lambda_4,$

$-\Lambda_1+\Lambda_2/\Lambda_1-\Lambda_4+\Lambda_5,\quad -\Lambda_1-\Lambda_2+\Lambda_4/\Lambda_1-\Lambda_5+\Lambda_6,\quad -\Lambda_1+\Lambda_3-\Lambda_4+\Lambda_5/\Lambda_1-\Lambda_6+\Lambda_7,$

$-\Lambda_1+\Lambda_3-\Lambda_5+\Lambda_6,\quad -\Lambda_3+\Lambda_5/\Lambda_1-\Lambda_7,\quad -\Lambda_1+\Lambda_3-\Lambda_6+\Lambda_7,$

$-\Lambda_3+\Lambda_4-\Lambda_5+\Lambda_6/-\Lambda_1+\Lambda_3-\Lambda_7,\quad -\Lambda_3+\Lambda_4-\Lambda_6+\Lambda_7,$

$\Lambda_2-\Lambda_4+\Lambda_6/-\Lambda_3+\Lambda_4-\Lambda_7,\quad \Lambda_2-\Lambda_4+\Lambda_5-\Lambda_6+\Lambda_7,$

$-\Lambda_2+\Lambda_6/\Lambda_2-\Lambda_4+\Lambda_5-\Lambda_7,\quad \Lambda_2-\Lambda_5+\Lambda_7,\quad -\Lambda_2+\Lambda_5-\Lambda_6+\Lambda_7/-\Lambda_2+\Lambda_5-\Lambda_7,$

$\Lambda_2-\Lambda_5+\Lambda_6-\Lambda_7,\quad -\Lambda_2+\Lambda_4-\Lambda_5+\Lambda_7/-\Lambda_2+\Lambda_4-\Lambda_5+\Lambda_6-\Lambda_7,\quad \Lambda_2-\Lambda_6,$

$\Lambda_3-\Lambda_4+\Lambda_7/-\Lambda_2+\Lambda_4-\Lambda_6,\quad \Lambda_3-\Lambda_4+\Lambda_6-\Lambda_7,\quad \Lambda_1-\Lambda_3+\Lambda_7/\Lambda_3-\Lambda_4+\Lambda_5-\Lambda_6,$

$\Lambda_1-\Lambda_3+\Lambda_6-\Lambda_7,\quad -\Lambda_1+\Lambda_7/\Lambda_3-\Lambda_5,\quad \Lambda_1-\Lambda_3+\Lambda_5-\Lambda_6,\quad -\Lambda_1+\Lambda_6-\Lambda_7/\Lambda_1-\Lambda_3+\Lambda_4-\Lambda_5,$

$-\Lambda_1+\Lambda_5-\Lambda_6/\Lambda_1+\Lambda_2-\Lambda_4,\quad -\Lambda_1+\Lambda_4-\Lambda_5/\Lambda_1-\Lambda_2,\quad -\Lambda_1+\Lambda_2+\Lambda_3-\Lambda_4/-\Lambda_1-\Lambda_2+\Lambda_3,$

$\Lambda_2-\Lambda_3/-\Lambda_2-\Lambda_3+\Lambda_4/-\Lambda_4+\Lambda_5/-\Lambda_5+\Lambda_6/-\Lambda_6+\Lambda_7/-\Lambda_7/.$

Using this list of weights and Corollary 4.3, we can write the following Kronecker products of the fundamental representations of the Lie algebra E_7:

$$(\Lambda_1)\otimes(\Lambda_7)=(\Lambda_1+\Lambda_7)\oplus(\Lambda_2)\oplus(\Lambda_7),\quad (\Lambda_2)\otimes(\Lambda_7)=(\Lambda_2+\Lambda_7)\oplus(\Lambda_3)\oplus(\Lambda_6)\oplus(\Lambda_1),$$
$$(\Lambda_3)\otimes(\Lambda_7)=(\Lambda_3+\Lambda_7)\oplus(\Lambda_1+\Lambda_2)\oplus(\Lambda_5)\oplus(\Lambda_1+\Lambda_7)\oplus(\Lambda_2),$$
$$(\Lambda_4)\otimes(\Lambda_7)=(\Lambda_4+\Lambda_7)\oplus(\Lambda_2+\Lambda_3)\oplus(\Lambda_1+\Lambda_5)\oplus(\Lambda_2+\Lambda_6)\oplus(\Lambda_3+\Lambda_7)\oplus(\Lambda_1+\Lambda_2)\oplus(\Lambda_5),$$
$$(\Lambda_5)\otimes(\Lambda_7)=(\Lambda_5+\Lambda_7)\oplus(\Lambda_4)\oplus(\Lambda_1+\Lambda_6)\oplus(\Lambda_2+\Lambda_7)\oplus(\Lambda_3)\oplus(\Lambda_6),$$
$$(\Lambda_6)\otimes(\Lambda_7)=(\Lambda_6+\Lambda_7)\oplus(\Lambda_5)\oplus(\Lambda_1+\Lambda_7)\oplus(\Lambda_2)\oplus(\Lambda_7),\quad (\Lambda_7)\otimes(\Lambda_7)=(2\Lambda_7)\oplus(\Lambda_6)\oplus(\Lambda_1)\oplus(\Lambda_0).$$

[1]For a partial list of references concerning the evidence for polynomial identities cf. Ref. 7.
[2]A. J. Bracken and H. S. Green, J. Math. Phys. **12**, 2099 (1971); H. S. Green, J. Math. Phys. **12**, 2106 (1971).
[3]K. C. Hannabuss, Mathematic Institute preprint, Oxford (1972).
[4]B. Kostant, J. Funct. Anal. **20**, 257 (1975).
[5]S. Okubo, J. Math. Phys. **18**, 2382 (1977).
[6]M. Iosifescu and H. Scutaru, in *Lecture Notes in Physics*, Vol. 313 (Springer, Berlin, 1988), p. 230.
[7]M. Iosifescu and H. Scutaru, J. Math. Phys. **29**, 742 (1988).
[8]N. Bourbaki, *Eléments de Mathématique, t.XXXVIII, Groupes et Algèbres de Lie* (Hermann, Paris, 1975), Chap. VIII.
[9]K. R. Parthasarathy, R. Ranga Rao, and V. S. Varadarajan, Ann. Math. **85**, 383 (1967).
[10]A. U. Klimyk, Ukrain. Mat. Z. **18**, 19 (1966).
[11]E. B. Dynkin, Trudy Mosk. Mat. Obsch. **1**, 39 (1952).
[12]A. J. Feingold, Proc. Am. Math. Soc. **70**, 109 (1978).
[13]E. B. Dynkin, Matem. Sbornik **30**, 349 (1952).
[14]F. Lemire and J. Patera, J. Math. Phys. **21**, 2026 (1980).
[15]A. Cavalli, G. D'Ariano, and L. Michel, Ann. Inst. H. Poincaré **44**, 173 (1986).
[16]S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1978).
[17]Y. Sakane and M. Takeuchi, J. Math. Soc. Jpn. **33**, 267 (1981).
[18]N. Yu. Reshetikhin and L. D. Faddeev, Teor. Mat. Fiz. **56**, 323 (1983); N. Yu. Reshetikhin, Teor. Mat. Fiz. **63**, 197, 347 (1985).
[19]H. Scutaru, J. Phys. A: Math. Gen. **20**, L1053 (1987).
[20]For early references cf. A. J. Macfarlane, L. O'Raifeartaigh, and P. S. Rao, J. Math. Phys. **8**, 536 (1967).
[21]E. B. Dynkin, Dokl. Akad. Nauk SSSR **71**, 221 (1950).
[22]M. D. Gould and S. A. Edwards, J. Phys. A: Math. Gen. **19**, 1523, 1531, 1537 (1986).
[23]F. Zaccaria, J. Math. Phys. **8**, 953 (1967).
[24]S. N. Kass, Commun. Alg. **15**, 2251 (1987).
[25]S. Kumar, Invent. Math. **93**, 117 (1988).

On left invariant Brownian motions and heat kernels of nilpotent Lie groups

Sergio Albeverio
Fakultät für Mathematik, Ruhr-Universität-Bochum, BiBoS-Bielefeld-Bochum, Federal Republic of Germany and SFB237 Essen-Bochum, Düsseldorf, Federal Republic of Germany

Teresa Arede
Faculdade da Engenharia, Universidade do Porto, Porto, Portugal

Zbigniew Haba
Institute of Theoretical Physics, University of Wroctaw, Wroctaw, Poland and SFB237 Essen-Bochum, Düsseldorf, Federal Republic of Germany

(Received 4 April 1989; accepted for publication 9 August 1989)

Left-invariant Brownian motions on nilpotent Lie groups are studied. Their characterization is given through Ito or Stratonovich stochastic differential equations, their generators are exhibited and the associated heat semigroups are studied. A reduction formula is given for these semigroups and their kernels, as integrals of products of normalized random Gaussian densities.

I. INTRODUCTION

The study of Brownian motion and diffusion processes on Lie groups has a long story. Already in 1928 Perrin studied Brownian motion on SO(3) and Brownian motion on U(1) was studied in details by P. Lévy in 1939.[1] Starting from the 1940s probability laws on (locally) compact groups were studied extensively, see, e.g., Refs. 2 and 3 and references therein. The systematic study of homogeneous processes on Lie groups was started by Hunt in 1956,[4] who used infinitesimal generator methods combined with probabilistic methods. Yosida[5] brought into this study new analytic tools.

Methods of stochastic differential equations were introduced in work by Ito[6] and by Mc Kean (see Ref. 7 and references therein), and much work has been done since, see, e.g., Refs. 8-21, and references therein.

For some applications of the study of stochastic processes on Lie groups in classical physics see, e.g., Refs. 19 and 20. The study of diffusion processes on Lie groups is also of importance in relation to certain problems of quantum mechanics, see, e.g., Refs. 21 and 22 and references therein.

In particular, motions on Lie groups arise directly in connection with spin particles, but also in a more indirect way in the study of models related to gauge fields (chiral models, σ models) and general relativity. A method for quantizing such motions is precisely to construct the heat kernel and the associated processes on the manifolds.

The study of the relations between quantum mechanics and classical mechanics also can be looked upon as the problem of controlling the heat kernel in the limit of small diffusion. Such problems, and related ones of small time asymptotics of heat kernels, have been studied quite extensively on a general level, see, e.g., Refs. 21, 23–29. Specific studies in certain groups, where the heat kernel is known explicitly, have also been done, see, e.g., Ref. 22 for compact groups and certain other homogeneous spaces (see also Refs. 30–32). Such studies are also of relevance in filter theory, see, e.g., Refs. 33 and 34 and references therein. In this paper we shall examine, somewhat with this point of view in mind, the case of nilpotent Lie groups.

Certain degenerate hypoelliptic differential operators on special Lie groups have been studied quite extensively as examples of hypoelliptic differential operators, see Refs. 25, 35–38. In our paper we shall examine the case of nondegenerate elliptic diffusion generators on general nilpotent Lie groups and obtain recursion resp. explicit formulas for the heat kernel.

In Sec. II we describe the natural left-invariant Brownian motion with drift on a Lie group, the associated generators, and semigroups.

In Sec. III we analyze the specific features of left-invariant Brownian motion with drift on a nilpotent Lie group and the properties of the associated semigroups. In particular we give a representation of the Laplacian on such groups and show the equivalence of the Ito and Stratonovich equations defining left-invariant Brownian motion on such groups.

In Sec. IV we study in more details the heat semigroup on nilpotent Lie groups and their kernels. In particular we give a recursion formula for reducing the computation of the heat semigroup on nilpotent Lie groups of order $n(n-1)/2$ to the one for such groups of order $[(n-1)(n-2)]/2$. We also exhibit the heat semigroup kernel as an integral of a product of normalized random Gaussian densities.

In Sec. V we study some examples and make some comments about the study of the small time behavior of the heat semigroup kernels.

II. LEFT-INVARIANT BROWNIAN MOTION ON A LIE GROUP AND ASSOCIATED SEMIGROUPS

Let us consider left-invariant Brownian motion (with drift) on an arbitrary Lie group G, following, e.g., Ref. 18, V. 35, p. 234. Let $v \in T_e G \cong g$ and let B be a Brownian motion on $\mathbb{R}^n$ and let B^m be its mth component. Let

$$Z \equiv B^m \tau_m + tv, \quad t \in \mathbb{R}_+, \tag{2.1}$$

with $\tau_m, m = 1,...,n$ a basis of g. Then Z is a (left-invariant) process (Brownian motion) on the additive Lie group $T_e G$. We make the convention of summing over repeated indices. A left-invariant Brownian motion $X(t)$, $t \in \mathbb{R}_+$ on G, with

 0022-2488/90/020278-09$03.00

drift given by v, is defined as the existent and unique solution of the stochastic differential equation in Stratonovich sense

$$\partial f(X(t)) = (L_m f)(X(t))\partial B^m + (L_v f)(X(t))\partial t, \quad (2.2)$$

for any $f \in C^\infty(G)$, with given $X(0)$, where L_v is the left-invariant vector field that takes the value v at the identity e of G; L_m is the unique left-invariant vector field associated with the element $\tau_m \in g$; and ∂ means Stratonovich differential. The above equation (2.2) is often written for short as

$$X^{-1}\partial X = \partial Z. \quad (2.3)$$

It is also said that X is obtained by product-integral injection of Z. (cf. also Ref. 7). For $v=0$ we have left-invariant Brownian motion on G without drift.

The characterizing properties of left-invariant Brownian motion (with drift) $X(t)$ on G are: (i) continuity of paths; (ii) $(X_s^{-1}X_{s+t}, t \geqslant 0)$ is independent of $\{X_r, r \leqslant s\} \forall s \geqslant 0$; (iii) for each $s \geqslant 0$ the processes $\{X_s^{-1}X_{s+t}, t \geqslant 0\}$ and $\{X_t, t \geqslant 0\}$ are identical in law. It is well known (by classical work of Yosida, Hunt, and others) that each Brownian motion on G is a Feller–Dynkin diffusion process (in particular strong Markov) on G. Its infinitesimal generator has the form

$$\mathscr{L} = \frac{1}{2}\sum_{m=1}^{n}(L_m)^2 + L_v. \quad (2.4)$$

The corresponding Feller–Dynkin semigroup $P_t \equiv e^{t\mathscr{L}}, t \geqslant 0$ has the property: P_t is a strongly continuous semigroup of linear operators on the space $C_0(G)$ of bounded continuous functions on G vanishing at infinity (looking upon G as a locally compact space) s.t., $0 \leqslant f \leqslant 1 \rightarrow 0 \leqslant P_t f \leqslant 1$, P_0 is the identity on $C_0(G)$ and $P_t f \rightarrow f$ as $t \downarrow 0$, $\forall f \in C_0(G)$, the convergence being in sup-norm.

Any such semigroup has a (normal) transition function $P_t(x,dy)$ s.t. $\forall t \geqslant 0$, $\forall x \in G$, $P_t(x,\cdot)$ is a measure on G (with its Borel structure), $P_t(x,G) \leqslant 1 \forall t \geqslant 0$, and $P_t(\cdot,\Gamma)$ is Borel measurable from G to $\mathbb{R}$; the Chapman–Kolmogorov equation holds; one has $P_0(x,\cdot) = \delta_x(\cdot)$ and

$$(P_t f)(x) = \int P_t(x,dy) f(y)$$

$\forall f$ bounded measurable on G.

Since $\{L_m, m = 1,\ldots,n\}$ is a basis of $g \cong T_x G \forall x \in G$, Hörmander's hypoellipticity condition (see, e.g., Ref. 18, V. 38, p. 253) is verified, hence $P_t(x,dy)$ has a density $p_t(x,y)$ with respect to Haar measure, for each $t>0$, $x \in G$, and $p(\cdot,\cdot,\cdot)$ is a C^∞ function on $(0,\infty) \times G \times G$. Moreover p satisfies Kolmogorov's forwards and backwards equations and

$$\lim_{t \downarrow 0} \int p(t,x,y) f(y) dy = f(x).$$

The transition semigroup P_t and transition density function $p_t(\cdot,\cdot,)$ are left-invariant in the sense that $P_t = e^{t\mathscr{L}}$ and $\mathscr{L}$ commutes on $C^\infty(G)$ with the generators L_m, $m = 1,\ldots,n$ of left translations.

Everything said until now concerning left-invariant Brownian motion and left-invariant transition semigroups can be repeated accordingly for right-invariant Brownian motion, defined by right-invariant vector fields and accordingly right-invariant transition semigroups.

Let us now introduce a metric on G. The canonical differential form (Maurer–Cartan form) ω on G is by definition the left differential of the identity map on G. It is thus a Lie-algebra-valued one-form and for each $x \in G$ we have the left invariance

$$\omega(x)h_x = x^{-1}h_x, \quad h_x \in T_x G \quad (2.5)$$

(with $x^{-1}h_x$ standing for the image of $h_x \in T_x G$ under the left translation induced by x^{-1}).

Let τ_i, $1 \leqslant i \leqslant n$, be a basis of the Lie algebra g. Then we can write

$$\omega(x)h_x = \sum_{i=1}^{n} \omega_h^i(x)\tau_i, \quad (2.6)$$

with $\omega_h^i(x) \in \mathbb{R}$. We have $\omega_h^i(x) = \langle h_x, \omega_i \rangle$, with $\omega^i \in (T_x G)^*$.

It is natural to introduce the following left invariant metric on G. A metric γ on G is defined naturally by giving on $T_x G$, $\forall x \in G$ the scalar product obtained by left translation from the scalar product on $T_e G \cong g$, which is the one coming from $\mathbb{R}^n$ (using the vector space structure of g as $\mathbb{R}^n$ and the Euclidean scalar product in $\mathbb{R}^n$).

Denoting by $\langle\ ,\ \rangle_x$ the scalar product at $x \in G$ we then have $\langle \tau_i, \tau_j \rangle_e = \delta_{ij}$ and for $h_x, h'_x \in T_x G$,

$$\begin{aligned}\langle h_x, h'_x \rangle_x &= \langle x^{-1}h_x, x^{-1}h'_x \rangle_e \\ &= \langle \omega(x)h_x, \omega(x)h'_x \rangle_e \\ &= \sum_{i=1}^{n} \omega_h^i, \omega_h^j \langle \tau_i, \tau_j \rangle_e \\ &= \sum_{i=1}^{n} \omega_h^i, \omega_h^i. \end{aligned} \quad (2.7)$$

This gives a metric γ on G which is by construction left invariant. (G,γ) is then a Riemannian space.

With respect to this Riemannian metric the Laplace-Beltrami operator is defined by

$$\Delta = (1/\sqrt{\gamma})\partial_i \gamma^{ij} \sqrt{\gamma}\, \partial_j, \quad (2.8)$$

with $\sqrt{\gamma} \equiv \sqrt{\det \gamma}$, with $\gamma^{ij}\gamma_{jk} = \delta_k^i$ and $\partial_i \equiv \partial / \partial x^i$ in local coordinates x, according to the general definition of Laplace–Beltrami operator on a Riemannian manifold. By the choice of the above left invariant metric we have

$$\langle L_i, L_j \rangle_\gamma = \delta_{ij}. \quad (2.9)$$

It follows then, see Ref. 18 (p. 217), that

$$\Delta = \sum_{m=1}^{n} (L_m)^2 - \Big(\sum_{m=1}^{n} k^r_{mm}\Big) L_r, \quad (2.10)$$

with

$$2k^i_{jk} \equiv c^i_{jk} + c^k_{ij} + c^j_{ik}, \quad (2.11)$$

c^i_{jk} being the structure constants such that

$$[L_j, L_k] = c^i_{jk} L_i. \quad (2.12)$$

The Brownian motion with drift on the manifold G with metric γ (as a stochastic process generated by the Laplace–Beltrami operator) coincides with the above left-invariant Brownian motion on G, if we take

$$v^q = -\frac{1}{2}\sum_m k^q_{mm}. \quad (2.13)$$

Remark: We can also express Δ in local coordinates $x=(x^1,...,x^n)$ on G by

$$\Delta=\gamma^{jk}(D_{jk}-\Gamma^i_{jk}D_i),$$

with

$$D_{jk}\equiv\frac{\partial}{\partial x^j}\frac{\partial}{\partial x^k},\quad D_i\equiv\frac{\partial}{\partial x^i},$$

$$\Gamma^i_{jk}\equiv\tfrac{1}{2}\gamma^{il}(D_j\gamma_{lk}+D_k\gamma_{lj}-D_l\gamma_{jk})$$

(the Christoffel symbols associated with the metric γ) (cf., e.g., Ref. 18, p. 216).

III. LEFT-INVARIANT BROWNIAN MOTION AND ITS GENERATOR ON NILPOTENT LIE GROUPS

Let g be a solvable Lie algebra over the complex resp. real numbers. By Ado's theorem there exists a faithful (i.e., injective) representation π of g in a (finite dimensional) complex vector space $V\neq\{0\}$ (e.g., the adjoint representation) [see, e.g., Ref. 39 (VI, § 2, p. 202]. It is possible to choose in V a basis such that all operators (endomorphisms) $\pi(l)$, $l\in g$ are given by upper triangular matrices, i.e., matrices A of the form $A_{ik}=0, i<k$ [see, e.g., Ref. 40 (2.3, Cor. 2, p. 402), Ref. 41 (III, Cor. 2.3)]. Nilpotent Lie algebras are special cases of solvable Lie algebras, and they have (finite dimensional) faithful representations in a complex vector space $V\neq\{0\}$ with a basis such that $\pi(l),l\in g$ are given by upper triangular matrices with elements on the diagonal being zero [this is part of Engel's theorem, see, e.g., Ref. 41, Chap. III, Th. 2.4 (p. 135)]. If G is a nilpotent connected Lie group, by using the above form of a faithful representation of its Lie algebra g together with the fact that the exponential mapping is a diffeomorphic (analytic) bijective map from g onto G, we see that on any nilpotent connected Lie group one can introduce a coordinate system (one single chart) such that any element $A\in G$ can be represented faithfully by a matrix of the form $A=((x^{ik}))$, $x_{ik}=0$ for $i<k$, $x^{ii}=1$ $\forall i=1,...,n$, if n is the dimension of G. To study nilpotent groups it is therefore sufficient to study the group of matrices of this form. Similarly, in the case of solvable Lie groups of exponential type, the exponential mapping is a surjective diffeomorphism, see, e.g., Ref. 42, p. 120. Hence we can parametrize also in this case any $A\in G$ in a single chart by matrices of the form $A=((x^{ik}))$ with $x^{ik}=0$ for $i<k$.

We shall now consider Brownian motion without drift on a nilpotent Lie group G. As we saw above it suffices to consider the case where G is a matrix group.

Lemma 3.1: On a nilpotent Lie group G the Laplace–Beltrami operator Δ defined in Sec. II is given by

$$\Delta=\partial_i\gamma^{ij}\partial_j.$$

Proof: From Sec. II we have that the left-invariant volume form on G is given by $\omega^1\wedge\omega^2\wedge...\wedge\omega^n$. By the definition of the metric on G by Lie exponentiation of the one given on g, we have that $\omega^1\wedge\omega^2\wedge...\wedge\omega^n$ is obtained by exponentiating to G the Riemann–Lebesgue volume element on g, the Jacobian determinant being

$$\det(d_y\exp),\quad y=\exp^{-1}x,\quad x\in G$$

(exp being the exponentiation from g to G).

By a known formula, see. e.g., Ref. 43 (XIX, 16, p. 219), we have

$$\det(d_y\exp)=\det\left[\sum_{p=0}^{\infty}\frac{1}{(p+1)!}(\mathrm{ad}(-x))^p\right].\tag{3.1}$$

Using the well-known formula

$$\det A=\exp\operatorname{tr}\ln A\tag{3.2}$$

we have that the rhs is equal to

$$\exp\operatorname{tr}\ln\sum_{p=0}^{\infty}\frac{1}{(p+1)!}(\mathrm{ad}(-x))^p.\tag{3.3}$$

But by using the representation of G by upper triangular matrices $\sum_{p=0}^{\infty}[1/(p+1)!](\mathrm{ad}(-x))^p$ is of the form $1+B$, where B is upper triangular with zero on the diagonal.

Using $\ln(1+B)=\sum_{n=1}^{\infty}[(-1)^{n+1}/n]B^n$ we get

$$\operatorname{tr}\ln(1+B)=0,\tag{3.4}$$

since $\operatorname{tr}B^n=0$. Introducing this into (3.1) and (3.2) we then get

$$\det(d_y\exp)=1,\tag{3.5}$$

which by the above yields that the volume element $\omega^1\wedge...\wedge\omega^n$ on G is the same as the one in the Lie algebra g.

By definition of the Haar measure on G we then have $\sqrt{\gamma}=1$. Introducng this into the general formula (2.7) for the Laplace–Beltrami operator Δ the lemma is proven. ■

From Sec. II we have the relation

$$\Delta=\sum_{m=1}^{n}(L_m)^2-\left(\sum_{m=1}^{n}k^r_{mm}\right)L_r\tag{3.6}$$

and Δ coincides with the infinitesimal generator of the left-invariant Brownian motion on G with drift $v^q=-\frac{1}{2}\Sigma_m k^q_{mm}$ In our case

$$\Delta=\partial_i\gamma^{ij}\partial_j.\tag{3.7}$$

Let us now consider the upper triangular matrix realization of G given above. A left-invariant vector field L_A on G corresponding to the element $A\in g$ is given by, with $x=((x^{ij}))\in G$, $i<j$, $i,j=1,...,N$ and $N(N-1)/2=n$

$$(L_Af)(x)=\sum_{i<j}(xA)^{ij}\frac{\partial}{\partial x^{ij}}f(x)\tag{3.8}$$

for any $f\in C^\infty(G)$.

In particular then, for the basis element ϵ_α of g in the direction $\alpha=lm, l<m$, [i.e., $(\epsilon_{lm})_{ij}=\delta_{il}\delta_{jm}$], denoting by L_{lm} the corresponding left-invariant vector field on G, we have

$$\begin{aligned}L_{lm}&=\sum_{i<j}(x\epsilon_{lm})^{ij}\frac{\partial}{\partial x^{ij}}\\&=\sum_{\substack{i<r\\i<j}}x^{ir}(\epsilon_{lm})_{rj}\frac{\partial}{\partial x^{ij}}\\&=\sum_{i<l}x^{il}\frac{\partial}{\partial x^{im}}.\end{aligned}\tag{3.9}$$

Introducing this into the expression (3.6) for Δ we get the representation

$$\Delta = \sum_{l<m}\Big(\sum_{i<l} x^{il}\frac{\partial}{\partial x^{im}}\Big)^2 - \sum_{r<s}\Big(\sum_{\alpha} k^{rs}_{\alpha\alpha}\Big)\sum_{i<r<s} x^{ir}\frac{\partial}{\partial x^{is}} \tag{3.10}$$

(with $k^{rs}_{\alpha\alpha}, r<s, 1\leqslant r<s\leqslant N$ denoting the set of numbers $k^{\beta}_{\alpha\alpha}, 1\leqslant r\leqslant n$ introduced before, now using the new basis in g as a matrix algebra).

Lemma 3.2:

$$\sum_{l<m}(L_{lm})^2 = \sum_{i,j,m}\Big(\sum_{l<m} x^{jl}x^{il}\Big)\frac{\partial^2}{\partial x^{jm}\partial x^{im}}$$
$$= \sum_{i,j,m}\frac{\partial}{\partial x^{jm}}\Big(\sum_{l} x^{jl}x^{il}\Big)\frac{\partial}{\partial x^{im}}.$$

Proof: From (3.9) we have, with $l<m$

$$(L_{lm})^2 = \Big(\sum_j x^{jl}\frac{\partial}{\partial x^{jm}}\Big)\Big(\sum_i x^{il}\frac{\partial}{\partial x^{im}}\Big)$$
$$= \sum_{i,j}\Big\{x^{jl}\delta_{ji}\delta_{ml}\frac{\partial}{\partial x^{im}} + x^{jl}x^{il}\frac{\partial^2}{\partial x^{jm}\,\partial x^{im}}\Big\}$$
$$= \sum_{i,j} x^{jl}x^{il}\frac{\partial^2}{\partial x^{jm}\,\partial x^{im}},$$

where in the last equality we used that $l<m$, hence $\delta_{ml}=0$. Summing now the above equality over m and l with $l<m$ we get the first equality in the lemma. The second equality follows by observing that by the same argument as above we have

$$(L_{lm})^2 = \sum_{i,j}\frac{\partial}{\partial x^{jm}}(x^{jl}x^{il})\frac{\partial}{\partial x^{im}}$$

and we then sum again over $l<m$. ■

Lemma 3.3: We have

$$v^{rs} = -\frac{1}{2}\sum_{l,m} k^{rs}_{lm,lm} = 0$$

(with the present realization of g as a matrix algebra and v, k defined accordingly as in Sec. II).

Proof: By definition of the structure constants we have, with $l<m$, $p<q$, $r<s$:

$$[L_{lm}, L_{pq}] = c^{rs}_{lm,pq}L_{rs}.$$

Introducing the expression (3.9) for L_{lm} we get then $c^{rs}_{lm,pq} = 0$ unless $r=l$, $s=q$ or $r=p$, $s=m$. In these cases we have $c^{lq}_{lm,pq} = \delta_{mp}$, $c^{pm}_{lm,pq} = -\delta_{ql}$ Inserting these values in the definition of v and $k^{rs}_{lm,lm}$ the lemma is proven. ■

Proposition 3.4: For the Laplace–Beltrami operator on the nilpotent group G with the metric described above and the given realization as matrix algebra we have:

$$\Delta = \sum_{\substack{m,l\\ l<m}}(L_{lm})^2 = \sum_{i,j,m,l} x^{jl}x^{il}\frac{\partial^2}{\partial x^{jm}\,\partial x^{im}}$$
$$= \sum_{i,j,m}\frac{\partial}{\partial x^{jm}}\Big(\sum_l x^{jl}x^{il}\Big)\frac{\partial}{\partial x^{im}}.$$

Proof: This is immediate from (3.6) and the above two lemmas. ■

From the metric given in Sec. II we have that the left invariant vector fields L_{lm} are mutually orthogonal, with our choice of the basis ϵ_{lm} in g. The matrix elements of the metric γ with respect to the base ϵ_{lm} are given by

$$\gamma_{ln,rm}(x) = \sum_{\substack{k<r\\ k<l}}(x^{-1})_{kl}(x^{-1})_{kr}\delta_{mn},$$

with $x\in G$. More precisely, these are the matrix elements of γ with respect to ϵ_{ln}, ϵ_{rm}. The inverse matrix to γ is given by

$$\gamma^{km,ln}(x) = \sum_r x^{kr}x^{lr}\delta_{mn}.$$

By the formula (3.7) for the Laplace–Beltrami operator on the nilpotent group G we have then, using this expression for γ^{-1}:

Proposition 3.5:

$$\Delta = \sum_{\substack{k<m\\ l<n}}\partial_{km}\gamma^{km,ln}\partial_{ln} = \sum_{\substack{k<m\\ l<m}}\partial_{km}\sum_r x^{kr}x^{lr}\partial_{lm}. \quad ■$$

Remark: This coincides with the expression for Δ given in Proposition 3.4. In deriving that proposition we used (3.6). The derivation of the latter proposition together with the two lemmas give an alternative direct proof, in our representation for G nilpotent, of (3.6).

The above results show in particular, that the Brownian motion with generator Δ in the case of nilpotent groups has zero drift. Hence, by the discussion of Sec. II, in such a case the Brownian motion with generator Δ is the solution of the Stratonovich equation (2.2), i.e.,

$$X^{-1}\,\partial X = \partial Z$$

with $Z = B^m\tau_m$ the Brownian motion (with zero drift) on the Lie algebra. This equation coincides with the Ito equation

$$dX = X\,dZ,$$

with d the Ito differential. This is so since the generator $\mathscr{L}$ of X has no first-order term, in this case, as follows from its expression (2.3), $v=0$, as computed above (Lemma 3.3 and Lemma 3.2). Hence we have proven the following theorem.

Theorem 3.6: Let G be a nilpotent Lie group, faithfully represented as a matrix group of upper triangular matrices. Then, the Laplace–Beltrami operator Δ on G is given as a sum of squares of vector fields:

$$\Delta = \sum_{l<m}(L_{lm})^2,$$

as given in Propositions 3.4 and 3.5.

Here, $\frac{1}{2}\Delta$ is the infinitesimal Markov generator of the stochastic process, left-invariant Brownian motion on G, given by the unique solution of the Stratonovich equation

$$X^{-1}\,\partial X = \partial B,$$

or equivalently the Ito equation

$$dX = X\,dB,$$

with B the Brownian motion on the Lie algebra.

Remark: A corresponding result concerning Δ on a compact Lie group (or a Lie group with unimodular Lie algebra) being given by a sum of squares of left-invariant vector fields was established by analytic methods in Ref. 44.

IV. THE HEAT SEMIGROUP AND ITS KERNEL ON NILPOTENT LIE GROUPS

Let G be a nilpotent Lie group and let $X(t)$, $t\in\mathbb{R}_+$ be the Brownian motion on G discussed in Sec. III. Let P_t be the corresponding Feller–Dynkin strong Markov diffusion semigroup discussed in Sec. III, i.e.,

$$(P_t f)(x) = E_x(f(X(t))), \tag{4.1}$$

for any $f\in C_0^\infty(G)$, where E_x is the expectation of X started at $x\in G$. Here, P_t has $\frac{1}{2}\Delta$, with Δ the Laplace–Beltrami operator on G, as its generator. We call P_t the heat semigroup on G. Let us now consider the matrix realization of G discussed in Sec. III. Then $x=((x^{ij}))$, with $x^{ij}=0$ for $i>j$, $x^{ii}=1$, $x^{ij}\in\mathbb{R}$, $1\leqslant i<j\leqslant N$, $N(N-1)/2\equiv n$. Using this global parametrization of G, f can be considered as a function on $\mathbb{R}^n$. Let us first study $P_t f$ for f of the following form.

Let $\mathscr{F}(\mathbb{R})$ be the set of functions h that are Fourier transforms of complex measures μ_h on $\mathbb{R}$, so that

$$h(\lambda) = \int_{\mathbb{R}} e^{i\lambda\alpha}\mu_h(d\alpha). \tag{4.2}$$

Here, $\mathscr{F}(\mathbb{R})$ is a Banach algebra with the norm

$$\|h\|_0 \equiv \|\mu_h\| \equiv \text{total variation of } \mu_h$$

see e.g., Ref. 45. We have that $\mathscr{F}(\mathbb{R})\cap L^p(\mathbb{R})$ is dense in all $L^p(\mathbb{R})$, $1\leqslant p<\infty$ and in $C_0(\mathbb{R})$ [since it contains, e.g., Schwartz test function space $\mathscr{S}(\mathbb{R})$]. Let $\mathscr{F}_n$ be the algebra of functions obtained by taking all linear combinations of functions of the form

$$h(\lambda_1,\dots,\lambda_n) \equiv \prod_{i=1}^n h_i(\lambda_i), \quad \text{with } h_i\in\mathscr{F}(\mathbb{R})\,\forall i.$$

Here, $\mathscr{F}_n$ is again a Banach algebra with norm

$$\|h\|_0 \equiv \|\mu_{h_1} * \dots * \mu_{h_n}\| \leqslant \prod_{i=1}^n \|\mu_{h_i}\|.$$

It is easily seen that $\mathscr{F}_n\cap L^p(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)\,\forall 1\leqslant p<\infty$ and in $C_0(\mathbb{R}^n)$.

For

$$f(x) = \prod_{1\leqslant i<j\leqslant N} f_{ij}(x^{ij})\in\mathscr{F}_n$$

we have

$$(P_t f)(x) = E_x(f(X(t))) = E_x\Big[\prod_{1\leqslant i<j\leqslant N} f_{ij}(X^{ij}(t))\Big]. \tag{4.3}$$

The Ito equation that determines X, given in Sec. III (Theorem 3.6)

$$dX = X\,dB$$

can be written, using the components X^{ij} of X and B^{ij} of B:

$$dX^{ij} = \sum_{l<j} X^{il}\,dB^{lj} \tag{4.4}$$

with $i<j$. In particular, using $X^{lm}=0$ for $l>m$, $X^{mm}=1$, $B^{NN}=0$ (since $B\in g$ and g is nilpotent):

$$X^{iN}(t) = X^{iN}(0) + \sum_{l=i}^{N-1}\int_0^t X^{il}\,dB^{lN}. \tag{4.5}$$

We shall consider the initial condition $X^{ij}(0)=x^{ij}$, with $x\in G$. Introducing this into (4.3), we get

$$\begin{aligned}(P_t f)(x) &= E_x\Big[\Big(\prod_{1\leqslant i<j\leqslant N-1} f_{ij}(X^{ij}(t))\Big)\Big(\prod_{1\leqslant i<j\leqslant N-1} f_{iN}(X^{iN}(t))\Big)\Big]\\ &= E_x\Big[\prod_{1\leqslant i<j\leqslant N-1} f_{ij}(X^{ij}(t))\int \exp\Big(i\sum_{j=1}^{N-1}\alpha_{jN}X^{jN}(t)\Big)\prod_{k=1}^{N-1} d\mu_{kN}(\alpha_{kN})\Big]\\ &= E_x\Big[\prod_{1\leqslant i<j\leqslant N-1} f_{ij}(X^{ij}(t))\int \exp\Big(i\sum_{j=1}^{N-1}\alpha_{jN}X^{jN}(0)\Big)\exp\Big(i\sum_{j=1}^{N-1}\alpha_{jN}\sum_{l=j}^{N-1}\int_0^t X^{jl}dB^{lN}(s)\Big)\prod_{k=1}^{N-1} d\mu_{kN}(\alpha_{kN})\Big]\end{aligned} \tag{4.6}$$

with $\mu_{iN}\equiv\mu_{f_{iN}}$. We take $X^{iN}(0)$ to be a.s. constant. By the Fubini theorem (4.6) is equal to

$$\int\prod_{k=1}^{N-1} d\mu_{kN}(\alpha_{kN})\exp\Big(i\sum_{j=1}\alpha_{jN}x^{jN}\Big)E_x\Big[\prod_{1\leqslant i<j\leqslant N-1} f_{ij}(X^{ij}(t))\exp\Big(i\sum_{j=1}^{N-1}\sum_{l=j}^{N-1}\alpha_{jN}\int_0^t X^{jl}\,dB^{lN}(s)\Big)\Big]. \tag{4.7}$$

By the independence of the Brownian motions B^{ij}, $i<j<N-1$ from B^{lN}, $l\leqslant N-1$ and the structure of the stochastic equations for $X(t)$, coming from the fact that $X(t)$ is upper triangular, the integrand depends on B^{lN}, $l=1,\dots N-1$, only through the terms involving $\int_0^t X^{il}\,dB^{lN}$, where X^{il} is independent of B^{lN}. Integrating with respect to the variables B^{lN}, $l\leqslant N-1$ by means of a well-known formula for Gaussian integration $E(\exp Z)=\exp[\frac{1}{2}E(Z^2)]$, i.e., in our case,

$$E\Big(\exp\Big(i\sum_{j=1}^{N-1}\alpha_{jN}\sum_{l=j}^{N-1}\int_0^t X^{jl}\,dB^{lN}\Big)\Big) = \exp\Big(-\frac{1}{2}\sum_{j=1}^{N-1}\alpha_{jN}\sum_{j'=1}^{N-1}\alpha_{j'N}\sum_{l=j\vee j}^{N-1}\int_0^t X^{jl}(s)X^{j'l}(s)ds\Big) \tag{4.8}$$

(with E expectation with respect to $B_{lN}, l\leqslant N-1$), we get

$$\begin{aligned}(P_t f)(x) &= \int\prod_{j=1}^{N-1} d\mu_{jN}(\alpha_{jN})\exp\Big(i\sum_{j=1}^{N-1}\alpha_{jN}x^{jN}\Big)\\ &\quad\times E_{N-1,x_{N-1}}\Big[\prod_{1\leqslant i<j\leqslant N-1} f_{ij}(X^{ij}(t))\exp\Big(-\frac{1}{2}\sum_{j,j'=1}^{N-1}\alpha_{jN}\alpha_{j'N}\sum_{l=j\vee j'}^{N-1}\int_0^t X^{jl}(s)X^{j'l}(s)ds\Big)\Big]\end{aligned} \tag{4.9}$$

Here we used the notation $E_{N-1,x_{N-1}}$ for the expectation with respect to the X^{ij}, $1\leqslant i<j\leqslant N-1$, with $x_{N-1}\equiv\{(x^{ij}), 1\leqslant i<j\leqslant N-1\}$.

Let us set

$$Y_{jj'}(t)\equiv\int_0^t\sum_{l=j\vee j'}^{N-1}X^{jl}(s)X^{j'l}(s)ds. \tag{4.10}$$

We can perform the integral with respect to the $\alpha_{jN}, j\leqslant N-1$ variables if we take $\mu_{iN}(d\alpha)$ to be an unnormalized complex Gaussian measure

$$\mu_{jN}(d\alpha)=\exp\left(-\eta\frac{|\alpha_{jN}|^2}{2}\right)e^{-i\alpha_{jN}x'_{jN}}d\alpha_{jN}, \tag{4.11}$$

for some $\eta>0$. In this case

$$(P_tf)(x)=E_{N-1,x_{N-1}}\left[\prod_{1\leqslant i<j\leqslant N-1}f_{ij}(X^{ij}(t))[\det(2\pi\widetilde{Y})]^{-1/2}\exp\left(-\frac{1}{2}\sum_{j,j'=1}^{N-1}(x_{jN}-x'_{jN})(\widetilde{Y}^{-1}(t))_{jj'}(x_{j'N}-x'_{j'N})\right)\right],$$

with $\widetilde{Y}_{jj'}\equiv Y_{jj'}+\eta\delta_{jj'}$. We remark that $\widetilde{Y}^{-1}$ exists since Y is a positive matrix and $\widetilde{Y}>Y$. Moreover $(\det 2\pi\widetilde{Y})^{-1/2}\leqslant(2\pi\eta)^{-(N-1)/2}$ so that the integrand is bounded and measurable for each $\eta>0$. The expression under the expection only depends on B^{ik}, $1\leqslant i<k\leqslant N-1$ (where we use again the upper triangular structure of the matrices entering the Ito equation $dX=X\,dB$).

Hence we have proven the following proposition.

Proposition 4.1: The semigroup P_t with generator $\frac{1}{2}\Delta$ on a nilpotent Lie group G parametrized by upper triangular matrices is given on functions f of the form $f(x)=\Pi_{1\leqslant i<j\leqslant(N-1)}f_{ij}(x^{ij})$, with $x=(x^{ij})$, $1\leqslant i,j\leqslant N, f_{ij}\in\mathscr{F}(\mathbb{R})$ such that f_{ij} is the Fourier transform of a complex measure μ_{ij}, by

$$(P_tf)(x)=\int_{\mathbb{R}^{N-1}}\prod_{j=1}^{N-1}d\mu_{jN}(\alpha_{jN})\exp\left(i\sum_{j=1}^{N-1}\alpha_{jN}x^{jN}\right)E_x\left[\left(\prod_{1\leqslant i<j\leqslant(N-1)}f_{ij}(X^{ij}(t))\right)\exp\left(i\sum_{j=1}^{N-1}\alpha_{jN}\sum_{l=1}^{N-1}\int_0^tX^{jl}dB^{lN}(s)\right)\right]$$

$$=\int_{\mathbb{R}^{N-1}}\prod_{j=1}^{N-1}d\mu_{jN}(\alpha_{jN})\exp\left(i\sum_{j=1}^{N-1}\alpha_{jN}x^{jN}\right)E_{N-1,x_{N-1}}\left[\prod_{1\leqslant i<j\leqslant N-1}f_{ij}(X^{ij}(t))\exp\left(-\frac{1}{2}\sum_{jj'=1}^{N-1}\alpha_{jN}\alpha_{j'N}Y_{jj'}(t)\right)\right],$$

where E_x is the expectation with respect to the Brownian motion X on G started at $x\in G$, given by

$$dX=X\,dB,$$

with B the Brownian motion on the Lie algebra g of G,

$$Y_{jj'}(t)\equiv\int_0^t\sum_{l=j\vee j'}^{N-1}X^{jl}(s)X^{j'l}(s)ds.$$

Here, $E_{N-1,x_{N-1}}$ is the expectation with respect to the $X^{ij}(t)$, $1\leqslant i<j\leqslant N-1$, x_{N-1} being x without the Nth row and the Nth column. If μ_{iN} is of the form

$$\mu_{jN}(d\alpha)=\exp\left(-\frac{\eta}{2}|\alpha_{jN}|^2\right)\times\exp(-i\alpha_{jN}x'^{jN})d\alpha_{jN},\quad \eta>0,$$

for some $x'\in G$, then

$$(P_tf)(x)=E_{N-1,x_{N-1}}\left[\prod_{1\leqslant i<j\leqslant N-1}f_{ij}(X^{ij}(t))[\det(2\pi\widetilde{Y})]^{-1/2}\times\exp\left(-\frac{1}{2}\sum_{j,j'=1}^{N-1}(x^{jN}-x'^{jN})(\widetilde{Y}^{-1}(t))_{jj'}\times(x^{j'N}-x'^{j'N})\right)\right]$$

with $\widetilde{Y}_{jj'}\equiv Y_{jj'}+\eta\delta_{jj'}$. ■

We find it convenient at this point to introduce the pinned Brownian motion (Brownian bridge) on the nilpotent group G_N of order N. Given any two points $a, b\in G_N$, the pinned Brownian motion conditioned to start at a at time 0 and to end at b at time t is by definition the probability measure $\nu_N^{(t,a,b)}$ defined for any $f\in C_b(\mathbb{R}^{nj}), j\in\mathbb{N}$, $t_1\leqslant t_2\leqslant\ldots\leqslant t_j$ by

$$\int d\nu_N^{(t,a,b)}f(X(t_1),\ldots,X(t_j))\equiv p_N(t,a,b)^{-1}\int p_N(t_1,a_1,x_1)p_N(t_2-t_1,x_1,x_2)\cdots p_N(t-t_j,x_j,b)f(x_1,\ldots,x_j)\prod_{k=1}^{j}dx_k, \tag{4.12}$$

with $p_N(\cdot,\cdot,\cdot)$ the transition probability density for the group G_N. Here, $\nu_N^{(t,a,b)}$ gives the transition probability kernel for the pinned Brownian motion $X(t)$.

In terms of this probability measure the formula for the heat semigroup P_t in the preceding proposition can be expressed as follows.

Proposition 4.2: Let G_N be a nilpotent group of order N. Let P_t^N be its heat semigroup as in Proposition 4.1. Let $\nu_{N-1}^{(\cdot,\cdot,\cdot)}$ be the transition probability for pinned Brownian motion on G_{N-1} and let $Y_{(N)}$ be defined by (4.10).

Then, for any $x=((x^{ij}), 1\leqslant i<j\leqslant N)\in G_N, [N(N-1)/2]=n$, we have

$$(P_t^Nf_N)(x)=(P_t^{N-1}f_{N-1}^*)(x),$$

where

$f^*_{N-1}(x^{*\prime},x)$

$$\equiv \int dx'\, d\nu_{N-1}^{(t,x^*,x^{*\prime})}(X(\cdot)) f_N(x')[\det(2\pi Y_{(N)})]^{-1/2} \times \exp[-\tfrac{1}{2}((\mathbf{x}-\mathbf{x}'),(Y_{(N)})^{-1}(t)(\mathbf{x}-\mathbf{x}'))],$$

with $x=(x^*,\mathbf{x})$, $\mathbf{x}=(x^{jN}, 1\leqslant j\leqslant N-1)$, (,) meaning scalar product in $\mathbb{R}^{N-1}$. Moreover, for f_N being the Fourier transform of the measure (4.11) we have

$$f^*_{N-1}(x^{*\prime},x)=f_{N-1}(x^{*\prime})h(x^{*\prime},x),$$

where

$$h(x^{*\prime},x)=\int dx'\, d\nu_{N-1}^{(t,x^*,x^{*\prime})}(X(\cdot))[\det(2\pi\widetilde{Y}_{(N)}]^{-1/2} \times\exp[-\tfrac{1}{2}((\mathbf{x}-\mathbf{x}'),(\widetilde{Y}_{(N)})^{-1}(t)(\mathbf{x}-\mathbf{x}'))]$$

Proof: The proof is an immediate consequence of Proposition 4.1 and the definition of ν_{N-1}. ■

Remark: Proposition 4.2 permits us to reduce the computation of the heat semigroup on G_N to the one on G_{N-1}.

From the discussion of Sec. II we know that P_t^N has a smooth kernel absolutely continuous with respect to Lebesgue measure on G_N, i.e.,

$$(P_t^N f_N)(x)=\int_{\mathbb{R}^n} p^N(t,x,y) f_N(y)dy; \tag{4.13}$$

$p^N(\cdot,\cdot,\cdot)$ is C^∞ on $(0,\infty)\times G_N\times G_N$.

Now let f_N be a function of the product form as in Proposition 4.1 and 4.2, i.e.,

$$f_N(x)\equiv f_N^\eta(x) \equiv \prod_{1\leqslant i<j\leqslant N-1} f_{ij}(x^{ij}) \prod_{j=1}^{N-1}\int_{\mathbb{R}} e^{i\alpha_{jN}x^{jN}} \times d\mu_{jN}{}^\eta(\alpha_{jN}), \tag{4.14}$$

with $d\mu_{jN}{}^\eta(\alpha_{jN})\equiv\exp[-(\eta/2)|\alpha_{jN}|^2-i\alpha_{jN}x'^{jN}]d\alpha_{jN}$.

Assuming the $f_{ij}(\cdot)$ are in $\mathscr{S}(\mathbb{R})$, $f_N^\eta(x)dx^*$ converges weakly as $\eta\searrow 0$ as a finite measure to

$$\prod_{1\leqslant i<j\leqslant N-1} f_{ij}(x^{ij})dx^* \prod_{j=1}^{N-1}\delta_{x'^{jN}}(x^{jN}) \tag{4.15}$$

Then, since $p^N(t,x,\cdot)$ is C^∞, $\int p^N(t,x,y) f_N^\eta(y)dy$ converges as $\eta\to 0$ to

$$\int p^N(t,x,(x'^*,\mathbf{x}'))\prod_{1\leqslant i<j\leqslant N-1} f_{ij}(x'_{ij})dx'^*, \tag{4.16}$$

with $(x'^*,\mathbf{x}')\equiv x'$.

From this and the result of Proposition 4.2 we get the following theorem.

Theorem 4.3: Let G_N, P_t^N be as in Proposition 4.2. Then the kernel of P_t^N has a smooth density $p^N(t,.,.)$ and we have for any x, $x'\in G_N$:

$$p^N(t,x,x') = p^{N-1}(t,x^*,x'^*) \times\int d\nu_{N-1}^{(t,x^*,x^{*\prime})}(X(\cdot))[\det(2\pi Y_{(N)}]^{-1/2} \times\exp[-\tfrac{1}{2}((\mathbf{x}-\mathbf{x}'),(Y_{(N)})^{-1}(t)(\mathbf{x}-\mathbf{x}'))],$$

where $x=(x^*,\mathbf{x})$, X is the pinned Brownian motion on G_{N-1}, $Y_{(N)}$ is defined in (4.10), and $\nu_{N-1}^{(t,x^*,x^{*\prime})}$ is defined by (4.12). ■

Remark: By repeating the procedure we get an expression for $p^N(t,x,x')$ as an integral of a product of normalized random Gaussian densities.

Remark: The integrand in Y_N is a principal $(N-1)\times(N-1)$ minor of the matrix $\gamma_{kr,lm}$ giving the metric on G_N (cf. Sec. III), evaluated at $X(s)$.

Heuristically for $|x-x'|$ small in $\mathbb{R}^n$ and $X(s)$ piecewise constant, the above observation together with Theorem 4.3 makes it plausible that the small diffusion limit of the heat semigroup should be given in terms of the geodesic distance between x and x', cf. the discussion in Sec. V.

V. SOME EXAMPLES OF HEAT SEMIGROUPS ON NILPOTENT GROUPS

Let us consider a nilpotent group G whose matrix representation is by matrices of the form

$$\begin{pmatrix}1 & x\\ 0 & 1\end{pmatrix},\quad x\in\mathbb{R}$$

(which corresponds to $N=2$, $n=1$). In this case the stochastic equation in Sec. III reduces to $dX(t)=dB(t)$, with B the Brownian motion on $\mathbb{R}$. Then of course $X(t)$ is simply Brownian motion on $\mathbb{R}$ and Theorem 4.3 reduces simply to

$$p(t,x,x')=(2\pi t)^{-1/2}\exp\left(\frac{-|x-x'|^2}{2t}\right).$$

In the case $N=n=3$ the group G_N consists of all 3×3 matrices of the form

$$x=\begin{pmatrix}1 & x_{12} & x_{13}\\ 0 & 1 & x_{23}\\ 0 & 0 & 1\end{pmatrix},\quad x_{ij}\in\mathbb{R},\ i<j.$$

This group of matrices is isomorphic to the classical Heisenberg group of dimension 3, whose Lie algebra is defined by the relations $[x,y]=-4z$, $[x,z]=[y,z]=0$. (See, e.g., Refs. 25 and 35.)

In Sec. III we got a realization of this algebra, as the one generated by L_{lm}, $1\leqslant l<m\leqslant 3$ with $L_{lm}=\Sigma_{i<l}x^{il}(\partial/\partial x^{im})$.

From Theorem 4.3 we have in this case for $t>0$

$$p(t,x,x')=(2\pi t)^{-1/2}\exp\left(\frac{|x_{12}-x'_{12}|^2}{2t}\right) \times\int d\nu^{(t,x_{12},x'_{12})}(X^{12}(t))[\det(2\pi Y)]^{-1/2} \times\exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{x}')Y^{-1}(\mathbf{x}-\mathbf{x}')\right], \tag{5.1}$$

with

$$Y\equiv\begin{pmatrix}t+\int_0^t X^{12}(s)X^{12}(s)ds & \int_0^t X^{12}(s)ds\\ \int_0^t X^{12}(s)ds & t\end{pmatrix}.$$

Here, $\nu^{(t,x_{12},x'_{12})}(X^{12}(t))$ is the probability measure giving the distribution of the pinned Brownian motion $X^{12}(t)$ starting at time 0 in x_{12}, and ending at time t in x'_{12}. This can be expressed by

$$X^{12}(s) = x^{12} + (x'_{12} - x_{12})(s/t) + \sqrt{t}\, q(s/t)\,, \quad (5.2)$$

where $q(s)$, $s \in [0,1]$ is the Brownian bridge with $q(0) = q(1) = 0$ (see, e.g., Ref. 29).

Inserting (5.2) into (5.1) we can replace integration with respect to $d\nu^{(t,x_{12},x'_{12})}$ by an integration with respect to the probability measure $d\mu_q(\cdot)$ associated with the Gaussian process with zero expectation and covariance

$$E(q(s)q(s')) = s(1-s'), \quad s < s'\,. \quad (5.3)$$

By a direct computation we get

$$\det 2\pi Y = (2\pi)^2 t^2 \left[1 + \frac{1}{t}\int_0^t (X^{12}(s))^2 ds - \frac{1}{t^2}\left(\int_0^t X^{12}(s) ds\right)^2\right] \quad (5.4)$$

and

$$Y^{-1} = (\det Y)^{-1}\begin{pmatrix} t & -\int_0^t X^{12}(s)ds \\ -\int_0^t X^{12}(s)ds & t + \int_0^t (X^{12}(s))^2 ds \end{pmatrix}. \quad (5.5)$$

Remark: The above representation (5.1) of the heat semigroups on the Heisenberg group is useful for the study of the kernel as $t \downarrow 0$. In fact from Schwarz inequality if follows from (5.4) that $\det Y \geqslant t^2$.

Moreover in (5.5) we have integrals involving $X^{12}(s)$ and by (5.2) this is a function of s/t and $\sqrt{t}\, q(s/t)$. We write as a shorthand X^{12} for $X^{12}(\sqrt{t}\, q(s/t), s/t)$. By changing the s-integration variables in Y into $s/t = s'$ we see that $p(t,x,x')$ get expressed in terms of $X^{12}(\sqrt{t}\, q(s'), s')$.

But $X^{12}(\sqrt{t}\, q(s'), s')$ depends linearly on $\sqrt{t}\, q(s')$ and inserting this into Y we see that the $\sqrt{t}$ dependence of Y, at fixed q in the probability space, is smooth for $t > 0$. The same considerations hold for Y^{-1} and $[\det(2\pi Y)]^{-1/2}$ (here we use also the above lower bound). Hence, we get that $(2\pi t)^{1/2}\exp(|x_{12} - x'_{12}|^2/2t)p(t,x,x')$ is given by an expectation with respect to the t-independent probability measure $\mu(q(\cdot))$ of an integrand that is smooth in $\sqrt{t}$ for almost all fixed paths, i.e., of the "Laplace form"

$$p(t,x,x') = (2\pi t)^{-3/2}\exp\left(-\frac{|x_{12} - x'_{12}|^2}{2t}\right)\cdot h(t)\,,$$

where

$$h(t) = \int \exp\left(-\frac{f}{t}(q(\cdot),\sqrt{t})\right) dq(\cdot)\,,$$

with $f(q(\omega),\sqrt{t})$ C^∞ in $\sqrt{t}$ for a.e. ω, with $h(\cdot) \in C^\infty(R_+)$.

The factor $t^{-3/2}$ comes from the $t^{-1/2}$ present in (5.1) combined with the t^2 in $\det 2\pi Y$. Hence, the study of the limit $t \downarrow 0$ of $p(t,x,x')$ is reduced to the study of a Laplace integral of the form

$$\int \exp\left(-\frac{f}{t}(q(\cdot),\sqrt{t})\right) dq(\cdot)\,.$$

To handle this integral one can use the Laplace method in function spaces (for such methods in general see, e.g., Refs. 25, 46, and 47).

This study of the $t \downarrow 0$ (or, equivalently, the "small diffusion limit") is based on the formula in Theorem 4.3 which holds for all n. So, in principle, the method can be extended to nilpotent groups of arbitrary order.

This should be confronted with studies of this problem for lower n's (e.g., Refs. 25 and 35) and with general (but somewhat implicit) results on short time expansions based on large deviations and Malliavin calculus, see, e.g., Ref. 48.

[1]P. Lévy, "L'addition des variables aléatoires définies sur une circonférence," Bull. Math. France 67 (1939); F. Perrin, "Etude mathématique du mouvement Brownien de rotation," Ann. Sci. Ec. Norm. Suppl. (1928).
[2]U. Grenander, *Probabilities on Algebraic Structures* (Wiley, New York, 1963).
[3]H. Heyer, *Probability Measures on Locally Compact Groups* (Springer, Berlin, 1977).
[4]G. A. Hunt, "Semi-groups of measures on Lie groups," Trans. Am. Math. Soc. **81**, 264–293 (1956).
[5]K. Yosida, "Brownian motion in a homogeneous Riemannian space," Pac. J. Math. **2**, 263–196 (1952).
[6]K. Ito, "Brownian motion on a Lie group," Proc. Jpn. Acad. **26**, 4–10 (1950).
[7]H. P. McKean, Jr., *Stochastic Integrals* (Academic, New York, 1969).
[8]R. Gangolli, "On the construction of certain diffusions on a differentiable manifold," Z. Wahrsch. verw. Geb. **2**, 209–419 (1964).
[9]S. A. Molchanov, "Diffusion processes and Riemannian geometry," Russ. Math. Surv. **30**, 1–53 (1975).
[10]M. Ibero, "Intégrales stochastiques multiplicatives et construction de diffusions sur un groupe de Lie," Bull. Soc. Math. France **100**, 175–191 (1976).
[11]M. Emery, "Stabilité des solutions des équations différentielles stochastiques: applications aux integrales multiplicatives stochastiques," Z. Warsch. theo. verw. Geb. **41**, 241–262 (1978).
[12]Ph. Feinsilver, "Processes with independent increments on a Lie group," Trans. Am. Math. Soc. **242**, 73–121 (1978).
[13]E. Jörgensen, "Construction of the Brownian motion and the Ornstein–Uhlenbeck process in a Riemannian manifold on the basis of the Gangolli–McKean injection scheme, Z. Wahrscheinlichkeitsth. verw. Geb. **44**, 71–87 (1978).
[14]H. Kunita, "On the representation of solutions of stochastic differential equations," Sém. Prob. XIV, Lect. Notes Math. **784**, 282–304 (1978/79).
[15]Y. Yamato, "Stochastic differential equations and nilpotent Lie algebras," Z. Wahrscheinlichk. vew. Geb. **47**, 213–229 (1979).
[16]M. Taylor, *Noncommutative Harmonic Analysis* (Am. Math. Soc., Providence, RI, 1986); R. L. Karandihar, "Girsanov type formula for a Lie group valued Brownian motion," in Sém. Prob. XVII 81/82, edited by J. Azema and M. Yor, *Lecture Notes in Mathematics*, Vol. 986 (Springer, Berlin, 1983), pp. 198–204.
[17]N. Th. Varopoulos, "Analysis on Lie groups," J. Funct. Anal. **76**, 346–410 (1988); R. L. Karandihar, *Multiplicative stochastic integration in Prediction Theory and Harmonic Analysis*, edited by V. Mandrekar and H. Salehi (North-Holland, Amsterdam, 1983), pp. 191–199.
[18]L. C. G. Rogers and D. Williams, *Diffusions, Markov Processes, and Martingales, Vol. 2. Itô Calculus* (Wiley, Chichester, 1987).
[19]I. Shigekawa, "Transformations of the Brownian Motion on the Lie Group," Taniguchi Symp. SA, Katata, 409–422 (1982); I. Shigekawa, Transformations of the Brownian Motion on a Riemannian Symmetric Space," Z. Wahrscheinlichkeitsth. verw. Geb. **65**, 493–522 (1984).
[20]M. Baldo, Brownian motion in group manifolds: application to spin relaxation on molecules," Physica **114**, 88–94 (1982).
[21]S. Albeverio and T. Arede, in: *Chaotic Behavior in Quantum Systems, Theory and Applications, Proceeding of Como Conference,* edited by G. Casati and J. Ford (Plenum, New York, 1984), pp. 37–76.
[22]T. Arede, La géométrie du noyau de la chaleur sur les variétés, Thèse 3ème cycle, Université d'Aix-Marseille II (Luminy), 1983; T. Arede, Manifolds for which the heat kernel is given in terms of geodesic lengths, Lett. Math. Phys. **9**, 121–131 (1985); T. Arede, Ph.D. thesis, Universidade classica de Lisboa (in preparation).

[23]S. Albeverio and R. Høegh-Krohn, "Oscillatory integrals and the method of stationary phase in infinitely many dimensions, with applications to the classical limit of quantum mechanics I, " Invent. Math. **40**, 59–106 (1977).
[24]S. Albeverio, P. Blanchard, and R. Høegh-Krohn, "Feynman Path Integrals and the Trace Formula for Schrödinger Operators," Commun. Math. Phys. **83**, 49–76 (1982).
[25]R. Azencott, P. Baldi, A. Bellaiche, C. Bellaiche, P. Bougerol, M. Chaleyat-Maurel, L. Elie, and J. Granara, "Géodésiques et diffusions en temps petit," Astérisque 84–85 (1981).
[26]J. Rezende, "The method of stationary phase for oscillatory integrals on Hilbert space," Commun. Math. Phys. **101**, 187–206 (1985).
[27]R. Léandre and G. Benarous, in *Stochastics, Algebra and Analysis in Classical and Quantum Dynamics*, Proc. Marseille Conf., Spring 88, edited by S. Albeverio, Ph. Blanchard, and D. Testard (Kluwer, Dordrecht, 1990).
[28]S. Kusuoka, "The general Malliavin Calculus based on Brownian sheet and Bismut's expansion for large deviation," in *Stochastic processes—Mathematics and Physics I*", edited by S. Albeverio, Ph. Blanchard, and L. Streit, Lecture Notes in Mathematics 1158 (Springer, Berlin, 1986), pp. 141–157, New York.
[29]B. Simon, *Functional Integration and Quantum Physics* (Academic, New York, 1979).
[30]K. D. Elworthy, A. Truman, Mnndumu: in *From Local Times to Global Geometry Control and Physics*, edited by K. D. Elworthy (Pitman, London, 1986), pp. 84–99.
[31]K. D. Elworthy, *Stochastic Differential Equations on Manifolds* (Cambridge U. P., Cambridge, 1982).
[32]K. D. Elworthy, in *Stochastic Profcesses—Physics and Geometry, Proceedings of the Ascona Conference 1988*, edited by S. Albeverio, G. Casati, U. Cattaneo, D. Merlini, and R. Moresi (World Scientific, Singapore, 1990).
[33]T. E. Duncan, "Stochastic systems in riemannian manifolds," J. Optim. Theor. Appl. **27**, 175–191 (1976).
[34]T. Arede, "A class of solvable nonlinear filters," Stochastics **23**, 377–389 (1988).
[35]B. Gaveau, "Principe de moindre action, propagation de la chaleur et estimées sous elliptiques sur certains groupes nilpotents," Acta Math. **139**, 95–153 (1977).
[36]A. Hulanicki, in *Probability Theory on Vector Spaces II*, edited by A. Weron, Lecture Notes in Mathematics, Vol. 828 (Springer, Berlin, 1980).
[37]N. K. Stanton, "The heat equation in several complex variables," Bull. Am. Math. Soc. **11**, 65–84 (1984).
[38]M. Pinsky and V. W. Wihstutz, "Lyapunov exponent of nilpotent Ito systems," Bremen Preprint, 1987.
[39]N. Jacobson, *Lie Algebras* (Wiley–Interscience, New York, 1962).
[40]M. Naimark, and A. Stern, *Théorie des representations des groupes* (Mir, Moscow, 1979).
[41]S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[42]A. Kirillov, Elements de la théorie des représentations (Mir, Moscow, 1974).
[43]J. Dieudonné, *Treatise on Analysis, Vol. IV* (Academic, New York, 1974).
[44]H. Urakawa, "On the least positive eigenvalue of the Laplacian for compact group manifolds," J. Math. Soc. Jpn. **31**, 209–226 (1979).
[45]S. Albeverio and R. Høegh-Krohn, *Mathematical Theory of Feynman Path Integrals*, Lecture Notes in Mathematics, Vol. 523 (Springer, Berlin, 1976).
[46]R. S. Ellis, *Entropy, Large Deviations and Statistical Mechanics* (Springer, New York, 1985).
[47]N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes* (North-Holland-Kodansha, Amsterdam, Tokyo, 1981).
[48]J. M. Bismut, *Large Deviations and Malliavin Calculus* (Birkhäuser, Basel, 1984).

Inhomogeneous boson realization of indecomposable representations of Lie algebras

Hong-Chen Fu and Chang-Pu Sun
CCAST (World Laboratory), P.O. Box 8730, Beijing, People's Republic of China and Department of Physics, Northeast Normal University, Changchun, Jilin Province, People's Republic of China[a]

(Received 5 April 1989; accepted for publication 29 August 1989)

By making use of the differential realization of Lie algebras in the space of inhomogeneous polynomials of a certain number of variables, the corresponding inhomogeneous boson realization of Lie algebras is given. A new kind of indecomposable representations of Lie algebras are studied on the universal enveloping algebra of Heisenberg–Weyl algebra, its subspaces and its quotient spaces. The finite-dimensional representations are naturally obtained on the subspaces of Fock space. As an example, the indecomposable and irreducible representations of the Lie algebra su(2) are discussed in detail.

I. INTRODUCTION

Indecomposable representations of physically relevant Lie algebras have been suggested for the description of unstable particles.[1,2] Gruber and his co-workers studied the indecomposable representations of Lie algebras on the universal enveloping algebra of this Lie algebra by making use of the purely algebraic method.[3,4] One of the authors has studied the indecomposable representations of Lie algebras on the universal enveloping algebra of Heisenberg–Weyl algebra, its subspaces and its quotient spaces by making use of the homogeneous boson realization (HBR) of Lie algebras.[5–7] In this paper we will study the indecomposable representations of Lie algebras by making use of the inhomogeneous boson realization (IHBR) of Lie algebras, which is obtained from the corresponding inhomogeneous differential realization (IHDR).

However, the IHDR of Lie algebras itself is very useful in "quasi-exactly-solvable problems of quantum mechanics" (QESP) discovered recently.[8–11] QESP have been proved to be related to the IHDR of Lie algebras. Turbiner has studied the one-dimensional QESP by using the IHDR of sl(2) algebra,[9] and pointed out that a similar procedure for the search of multidimensional QESP can be developed if we use the IHDR of sl(m) algebra.[9] Shifman and Turbiner studied the two-dimensional QESP by making use of the IHDR of su(2)×su(2), so(3), and su(3) algebras.[10] Therefore, the IHDR of Lie algebras given in this paper will play an important role in the search for multidimensional QESP.

In this paper, the IHDR of Lie algebras is given by generalizing Shifman's discussions in Ref. 11. The corresponding IHBR of Lie algebras is obtained with provision for the corresponding relation between differential operators and creation and annihilation operators of boson states. In comparison with the HBR given in Refs. 5–7, the IHBR of Lie algebras uses creation and annihilation operators less than the HBR, and enables us to obtain the finite-dimensional representations on the subspaces of Fock space, while in Refs. 5–7 we can only obtain the finite-dimensional representations on the quotient spaces of Fock space. By making use of the IHBR of Lie algebras, a new class of indecomposable representations of Lie algebras are studied on the universal enveloping algebra of Heisenberg–Weyl algebra, its subspaces and its quotient spaces. As an explicit example, the indecomposable representations of su(2) algebra are studied in detail.

Although the IHDR of Lie algebras given in this paper looks trivial, we can use its corresponding IHBR to obtain various nontrivial indecomposable representations on the universal enveloping algebra of Heisenberg–Weyl algebra, its subspaces and its quotient spaces.

The symbols $\mathbb{Z}^+$ and $\mathbb{N}$ denote the set of non-negative integers and the set of positive integers, respectively. The symbol $\mathbb{C}$ denotes the field of complex numbers.

II. FROM IHDR TO IHBR

A. IHDR of Lie algebras

Let the basis for the Lie algebra L be all $\{T_p\}$ that satisfy the Lie product $[T_p,T_q] = \Sigma_r C^r_{pq} T_r$, where C^r_{pq} are constructure constants. Let D be a faithful representation of L with dimension $m < \infty$, and let $\{e_1,e_2,...,e_m\}$ be the basis for a representation space. Then we have

$$T_p e_i = \sum_{j=1}^{m} D(T_p)_{ji} e_j \, . \qquad (2.1)$$

Since we would like to construct a realization on polynomials, we introduce m independent variables $\{x_1,x_2,...,x_m\}$ and identify them with the basis vectors $\{e_1,e_2,...,e_m\}$:

$$x_i \Leftrightarrow e_i \quad (i = 1,2,...,m) \, . \qquad (2.2)$$

Now Eqs. (2.1) and (2.2) imply

$$T_p x_i = \sum_{j=1}^{m} D(T_p)_{ji} x_j \, . \qquad (2.3)$$

Equation (2.3) immediately allows us to write the T_p in the differential form

$$T_p = \sum_{ij=1}^{m} D(T_p)_{ji} x_j \frac{\partial}{\partial x_i} \, . \qquad (2.4)$$

This realization is obviously valid not only for the first-order homogeneous polynomial space with basis $\{x_1,x_2,...,x_m\}$ that carries the representation D but also for the space of nth-order homogeneous polynomials spanned by

[a] Mailing address.

 0022-2488/90/020287-07$03.00

$$\left\{x_1^{i_1}x_2^{i_2}\cdots x_m^{i_m}\,\Bigg|\,\sum_{k=1}^{m} i_k = n,\ i_k\in\mathbf{Z}^+\right\}, \tag{2.5}$$

which carries the symmetrized direct product representation

$$D_s^{\otimes n} = \underbrace{(D\otimes D\otimes\cdots\otimes D)}_{n\text{ copies}}\ \text{symmetrized}. \tag{2.6}$$

In fact, the following equation defines a new realization $\hat{T}$ on the space of nth-order homogeneous polynomials:

$$\hat{T}_p(x_1^{i_1}x_2^{i_2}\cdots x_m^{i_m}) = \sum_{r=1}^{m} x_1^{i_1}\cdots x_{r-1}^{i_{r-1}}(T_p x_r^{i_r})x_{r+1}^{i_{r+1}}\cdots x_m^{i_m}. \tag{2.7}$$

From Eqs. (2.4) and (2.7) it follows that

$$\hat{T}_p = \sum_{r,j=1}^{m} D(T_p)_{jr}x_j\frac{\partial}{\partial x_r} \equiv T_p. \tag{2.8}$$

We call the realization (2.8) on the space of homogeneous polynomials the homogeneous differential realization (HDR) that corresponds to the HBR given in Refs. 5–7.

However, we would like to construct the differential realization on the space of inhomogeneous polynomials for the needs of QESP. For this purpose we define a new variable:

$$\xi_i = x_i/x_m \quad (i=1,2,\ldots,m-1), \tag{2.9}$$

$$x_1^{i_1}x_2^{i_2}\cdots x_m^{i_m}$$

$$\Rightarrow \xi_1^{i_1}\xi_2^{i_2}\cdots\xi_{m-1}^{i_{m-1}}x_m^n \quad \left(\sum_{k=1}^{m-1} i_k = 0,1,2,\ldots,n\right). \tag{2.10}$$

Then the linear space $\mathscr{P}$ spanned by

$$\mathscr{P}:\left\{\xi_1^{i_1}\xi_2^{i_2}\cdots\xi_{m-1}^{i_{m-1}}x_m^n\,\Bigg|\,\sum_{k=1}^{m-1} i_k = 0,1,\ldots,n,\ i_k\in\mathbf{Z}^+\right\} \tag{2.11}$$

is a space of inhomogeneous polynomials with dimension

$$\dim\mathscr{P} = \sum_{k=0}^{n}\frac{(m+k-2)!}{k!\,(m-2)!}. \tag{2.12}$$

In fact, we can regard $\{\xi_i | i=1,2,\ldots,m-1\}$ as the local coordinates of the projective space

$$PR^{m-1}: \{[x] | [x] = \{y\in\mathbf{R}^n | y=\lambda x,\ \lambda\in\mathbf{R}\}\}$$

of

$$\mathbf{R}^n: \{x=[x_1,\ldots,x_m] | x_1,\ldots,x_m \in \mathbf{R}\},$$

where $\mathbf{R}$ is the field of real numbers. The space $\mathscr{P}$ is the polynomial space with regard to the local coordinates $\{\xi_i | i=1,2,\ldots,m-1\}$ of PR^{m-1}, one of whose elements corresponds to a polynomial in $\mathbf{R}^n$ through the corresponding relation (2.10) between the basis for $\mathscr{P}$ and the basis for the space of homogeneous polynomials.

It is easy to deduce that

$$\hat{T}_p x_m^n = n x_m^n D(T_p)_{mm} + n x_m^n \sum_{i=0}^{m-1} D(T_p)_{im}\xi_i,$$

$$\hat{T}_p\,\xi_k^{i_k} = D(T_p)_{mk} i_k\,\xi_k^{i_k-1} + \sum_{i=1}^{m-1} D(T_p)_{ik} i_k \xi_i\,\xi_k^{i_k-1} - D(T_p)_{mm} i_k\,\xi_k^{i_k} - \sum_{i=1}^{m-1} D(T_p)_{im} i_k\,\xi_i\,\xi_k^{i_k}.$$

Therefore

$$\begin{aligned}
\hat{T}_p(\xi_1^{i_1}\xi_2^{i_2}\cdots\xi_{m-1}^{i_{m-1}}x_m^n) &= nD(T_p)_{mm}(\xi_1^{i_1}\xi_2^{i_2}\cdots\xi_{m-1}^{i_{m-1}}x_m^n) + n\sum_{i=1}^{m-1} D(T_p)_{im}\xi_i\,\xi_1^{i_1}\xi_2^{i_2}\cdots\xi_{m-1}^{i_{m-1}}x_m^n \\
&\quad + \sum_{k=1}^{m-1} D(T_p)_{mk}\frac{\partial}{\partial\xi_k}(\xi_1^{i_1}\xi_2^{i_2}\cdots\xi_{m-1}^{i_{m-1}}x_m^n) + \sum_{k=1}^{m-1}\sum_{i=1}^{m-1} D(T_p)_{ik}\,\xi_i\frac{\partial}{\partial\xi_k} \\
&\quad \times(\xi_1^{i_1}\xi_2^{i_2}\cdots\xi_{m-1}^{i_{m-1}}x_m^n) - \sum_{k=1}^{m-1} D(T_p)_{mm}\,\xi_k\frac{\partial}{\partial\xi_k}(\xi_1^{i_1}\xi_2^{i_2}\cdots\xi_{m-1}^{i_{m-1}}x_m^n) \\
&\quad - \sum_{k=1}^{m-1}\sum_{i=1}^{m-1} D(T_p)_{im}\,\xi_i\,\xi_k\frac{\partial}{\partial\xi_k}(\xi_1^{i_1}\xi_2^{i_2}\cdots\xi_{m-1}^{i_{m-1}}x_m^n).
\end{aligned} \tag{2.13}$$

From (2.13) we obtain the desired IHDR on $\mathscr{P}$:

$$\begin{aligned}
\hat{T}_p &= nD(T_p)_{mm} + n\sum_{i=1}^{m} D(T_p)_{im}\,\xi_i + \sum_{i=1}^{m-1}\sum_{k=1}^{m-1} D(T_p)_{ik}\,\xi_i\frac{\partial}{\partial\xi_k} + \sum_{k=1}^{m-1} D(T_p)_{mk}\frac{\partial}{\partial\xi_k} \\
&\quad - D(T_p)_{mm}\sum_{k=1}^{m-1}\xi_k\frac{\partial}{\partial\xi_k} - \sum_{i=1}^{m-1}\sum_{k=1}^{m-1} D(T_p)_{im}\,\xi_i\,\xi_k\frac{\partial}{\partial\xi_k}.
\end{aligned} \tag{2.14}$$

It is easy to check the commutation relation

$$[\hat{T}_p, \hat{T}_q] = \sum_r C_{pq}^r\,\hat{T}_r. \tag{2.15}$$

It should be noted that n in the basis for $\mathscr{P}$ is a positive integer, and that the x_m^n is an overall factor. Its existence provides the first and second terms in the realization of (2.14). However, if we extend the positive integer n to an arbitrary real number, we can also check that the realization (2.14) with real number n also satisfies the Lie product

(2.15), e.g., it is also the differential realization of L. Therefore the positive integer n in (2.14) can be extended to an arbitrary real number.

In comparison to the more difficult situation presented in Appendix A of Ref. 11 with su(3) as an example, the IHDR obtained in this paper looks very trivial because it only covers those representations that are the symmetrized direct product of one fundamental representation [e.g., a triplet or an antitriplet for su(3)] and are marked by one positive integer n. It is well known that the finite-dimensional (irreducible) representations of semisimple Lie algebras are marked by rank L non-negative integers, where rank L is the rank of semisimple Lie algebra L. In order to obtain all representations of semisimple Lie algebras in the space of polynomials, all the fundamental representations must be exhausted. In the product of all the fundamental representations we impose certain additional conditions and then obtain the nontrivial IHDR marked by rank L integers, each one of which is the number of a fundamental representation in the product. However, it is difficult to obtain analytically such an IHDR in practice. Because the main purpose of this paper is to obtain the indecomposable representations in a differential way, we only need the IHBR that corresponds to the most trivial IHDR (2.14). In the following discussions we will see that the indecomposable representations are marked by a certain number of complex numbers involving n on the quotient spaces of the universal enveloping algebra of Heisenberg–Weyl algebra.

B. IHBR of Lie algebras

Notice the corresponding relation between creation and annihilation operators of $(m-1)$-boson states $\{a_i^+, a_i \mid i=1,2,\dots,m-1\}$ and the operators $\{\xi_i, \partial/\partial\xi_i \mid i=1,2,\dots,m-1\}$ in $\mathscr{P}$,

$$a_i^+ \Leftrightarrow \xi_i, \quad a_i \Leftrightarrow \partial/\partial\xi_i, \tag{2.16}$$

and the same commutation relations:

$$[a_i,a_j^+] = \delta_{ij}E,$$
$$[E,a_i] = [E,a_i^+] = 0 \quad (E = \text{identity operator}), \tag{2.17}$$

$$\left[\xi_i, \frac{\partial}{\partial\xi_j}\right] = \delta_{ij}1, \quad [1,\xi_i] = \left[1, \frac{\partial}{\partial\xi_i}\right] = 0.$$

We obtain the IHBR of Lie algebras from (2.14):

$$\begin{aligned}B(T_p) = {} & nD(T_p)_{mm} + n\sum_{i=1}^{m-1} D(T_p)_{im}a_i^+ \\ & + \sum_{k=1}^{m-1} D(T_p)_{mk}a_k \\ & + \sum_{i=1}^{m-1}\sum_{k=1}^{m-1} D(T_p)_{ik}a_i^+a_k \\ & - D(T_p)_{mm}\sum_{k=1}^{m-1} a_k^+a_k \\ & - \sum_{i=1}^{m-1}\sum_{k=1}^{m-1} D(T_p)_{im}a_i^+a_k^+a_k.\end{aligned} \tag{2.18}$$

When we regard E as 1 we have

$$[B(T_p),B(T_q)] = \sum_r C_{pq}^r B(T_r). \tag{2.19}$$

Although the IHDR and the IHBR of Lie algebras satisfy the same commutation relations, the IHBR can give richer representations than the IHDR. In fact, $\mathscr{P}$ is isomorphous to the subspace $\mathscr{F}(n)$, spanned by

$$\{a_1^{+i_1}a_2^{+i_2}\cdots a_{m-1}^{+i_{m-1}}|0\rangle \mid i_1,i_2,\dots,i_{m-1}\in\mathbb{Z}^+,$$
$$i_1+i_2+\cdots+i_{m-1} = 0,1,2,\dots,n, \quad a_k|0\rangle = 0\},$$

of the Fock space $\mathscr{F}$ with basis

$$\mathscr{F}: \{a_1^{+i_1}a_2^{+i_2}\cdots a_{m-1}^{+i_{m-1}}|0\rangle \mid i_1,i_2,\dots,i_{m-1}\in\mathbb{Z}^+,$$
$$a_k|0\rangle = 0\}.$$

The Fock space $\mathscr{F}$ is isomorphic to the quotient space $\mathscr{F}' \equiv (\Omega/I)/J$ of the universal enveloping algebra Ω of the $(m-1)$-state Heisenberg–Weyl algebra $\mathscr{H}$: $\{a_i^+,a_i,E \mid i = 1,2,\dots,m-1\}$ with the PBW basis

$$\Omega: \left\{ X(r_i,s_i,t) \equiv E^t\left(\prod_{i=1}^{m-1} a_i^{+r_i}\prod_{i=1}^{m-1} a_i^{s_i}\right) \middle| \, r_i,s_i,t\in\mathbb{Z}^+\right\}, \tag{2.20}$$

where I is a left ideal generated by $(E-1)$, and J is a left ideal generated by $\{a_i \mid i=1,2,\dots,m-1\}$. The space $V\equiv\Omega/I$ with basis

$$V: \{X(r_i,s_i) \equiv X(r_i,s_i,0) \bmod I \mid r_i,s_i\in\mathbb{Z}^+\}, \tag{2.21}$$

which carries the representation $\rho[\rho(E) = 1]$ of $\mathscr{H}$, is larger than $\mathscr{F}'$. Therefore the representations of $B(T_p)$ on V are richer than the representations of $\hat{T}_p$ on $\mathscr{P}$. This is why we study the representations by making use of the IHBR, instead of the IHDR.

Comparing with the homogeneous boson realization given in Refs. 5 and 6, the IHBR has merit: It only uses creation and annihilation operators of $(m-1)$-boson states for the m-dimensional faithful representation D of Lie algebra L, while the homogeneous boson realization must use creation and annihilation operators of m-boson states.

C. Example: IHDR and IHBR of sl(m) algebra

We choose the basis for sl(m) algebra as

$$\begin{aligned}\{ & T_{ij} = e_{ij} \quad (i\neq j = 1,2,\dots,m), \\ & T_k = e_{kk} - e_{(k+1)(k+1)}, \\ & (k = 1,2,\dots,m-1)\},\end{aligned} \tag{2.22}$$

where e_{ij} is a $m\times m$ matrix with matrix element $(e_{ij})_{pq} = \delta_{ip}\delta_{jq}$. The finite-dimensional representation D of sl(m) algebra is chosen as the natural representation, e.g., $D(T_{ij}) = T_{ij}$, $D(T_k) = T_k$. By making use of Eq. (2.14) we obtain the IHDR of sl(m) algebra:

$$\begin{aligned}& \hat{T}_{pq} = \xi_p\frac{\partial}{\partial\xi_q} \quad (p\neq q = 1,2,\dots,m-1), \\ & \hat{T}_{pm} = n\xi_p - \xi_p\sum_{j=1}^{m-1}\xi_j\frac{\partial}{\partial\xi_j} \quad (p = 1,2,\dots,m-1), \\ & \hat{T}_{mq} = \frac{\partial}{\partial\xi_q} \quad (q = 1,2,\dots,m-1), \\ & \hat{T}_p = \xi_p\frac{\partial}{\partial\xi_p} - \xi_{p+1}\frac{\partial}{\partial\xi_{p+1}} \quad (p = 1,2,\dots,m-2),\end{aligned} \tag{2.23}$$

$$\hat{T}_{m-1} = \xi_{m-1}\frac{\partial}{\partial \xi_{m-1}} - n + \sum_{j=1}^{m-1} \xi_j \frac{\partial}{\partial \xi_j}.$$

The corresponding IHBR is

$$\begin{aligned}
&B(T_{pq}) = a_p^+ a_q \quad (p \neq q = 1,2,\ldots, m-1),\\
&B(T_{pm}) = na_p - a_p \sum_{j=1}^{m-1} a_j^+ a_j \quad (p = 1,2,\ldots, m-1),\\
&B(T_{mq}) = a_q \quad (q = 1,2,\ldots, m-1), \qquad (2.24)\\
&B(T_p) = a_p^+ a_p - a_{p+1}^+ a_{p+1} \quad (p = 1,2,\ldots, m-2),\\
&B(T_{m-1}) = a_{m-1}^+ a_{m-1} - n + \sum_{j=1}^{m-1} a_j^+ a_j.
\end{aligned}$$

When $m = 2$ we obtain the IHDR and IHBR of sl(2) algebra:

$$\hat{T}^+ = n\xi - \xi^2 \frac{d}{d\xi}, \quad \hat{T}^- = \frac{d}{d\xi}, \quad \hat{T}^0 = -\frac{1}{2}n + \xi\frac{d}{d\xi}, \qquad (2.25a)$$

$$B(T^+) = na^+ - a^{+2}a, \quad B(T^-) = a, \quad B(T^0) = -\tfrac{1}{2}n + a^+a, \qquad (2.25b)$$

where $T^+ = T_{12}$, $T^- = T_{21}$, and $T^0 = \frac{1}{2}T_1$. Equations (2.25a) are just the realization adopted by Refs. 8 and 9.

III. INDECOMPOSABLE REPRESENTATIONS OF LIE ALGEBRAS

The representation of the Heisenberg–Weyl algebra $\mathscr{H}$ on its universal enveloping algebra Ω is defined

$$\begin{aligned}
&\rho(a_j^+)X(r_i,s_i,t) = X(r_i + \delta_{ij},s_i,t),\\
&\rho(a_k)X(r_i,s_i,t) = X(r_i,s_i + \delta_{ik},t) + r_k X(r_i - \delta_{ik},s_i,t+1), \qquad (3.1)\\
&\rho(E)X(r_i,s_i,t) = X(r_i,s_i,t+1).
\end{aligned}$$

It induces a representation of $\mathscr{H}$ on $V = \Omega/I$,

$$\begin{aligned}
&\rho(a_j^+)X(r_i,s_i) = X(r_i + \delta_{ij},s_i),\\
&\rho(a_k)X(r_i,s_i) = X(r_i,s_i + \delta_{ik}) + r_k X(r_i - \delta_{ik},s_i), \qquad (3.2)\\
&\rho(E)X(r_i,s_i) = X(r_i,s_i) \quad [\text{e.g.}, \rho(E) = 1].
\end{aligned}$$

By making use of the equation

$$\begin{aligned}
\Gamma(T_p) = {}& nD(T_p)_{mm} + n\sum_{i=1}^{m-1} D(T_p)_{im}\,\rho(a_i^+) + \sum_{k=1}^{m-1} D(T_p)_{mk}\,\rho(a_k) + \sum_{i=1}^{m-1}\sum_{k=1}^{m-1} D(T_p)_{ik}\,\rho(a_i^+)\,\rho(a_k)\\
&- D(T_p)_{mm}\sum_{k=1}^{m-1}\rho(a_k^+)\,\rho(a_k) - \sum_{i=1}^{m-1}\sum_{k=1}^{m-1} D(T_p)_{im}\,\rho(a_i^+)\,\rho(a_k^+)\,\rho(a_k), \qquad (3.3)
\end{aligned}$$

we obtain the representation Γ of the Lie algebra L on V:

$$\begin{aligned}
\Gamma(T_p)X(r_i,s_i) = {}& \left(n - \sum_{k=1}^{m-1} r_k\right)\sum_{j=1}^{m-1} D(T_p)_{jm}X(r_i + \delta_{ij},s_i) + \left(n - \sum_{k=1}^{m-1} r_k\right) D(T_p)_{mn}X(r_i,s_i)\\
&- \sum_{j=1}^{m-1}\sum_{k=1}^{m-1} D(T_p)_{jm}X(r_i + \delta_{ik} + \delta_{ij},s_i + \delta_{ik}) + \sum_{j=1}^{m-1}\sum_{k=1}^{m-1} D(T_p)_{jk}X(r_i + \delta_{ij},s_i + \delta_{ik})\\
&- D(T_p)_{mm}\sum_{k=1}^{m-1} X(r_i + \delta_{ik},s_i + \delta_{ik}) + \sum_{k=1}^{m-1} D(T_p)_{mk}X(r_i,s_i + \delta_{ik})\\
&+ \sum_{j=1}^{m-1}\sum_{k=1}^{m-1} D(T_p)_{jk} r_k X(r_i - \delta_{ik} + \delta_{ij},s_i) + \sum_{k=1}^{m-1} D(T_p)_{mk} r_k X(r_i - \delta_{ik},s_i). \qquad (3.4)
\end{aligned}$$

It is observed that the value $\Sigma_{i=1}^{m-1} s_i$ cannot decrease in Eq. (3.4). Thus each non-negative integer N defines a Γ-invariant subspace $V^{[N]}$ of V with basis

$$V^{[N]}: \left\{ X(r_i,s_i) \,\middle|\, \sum_{i=1}^{m-1} s_i \geqslant N,\ r_i,s_i \in \mathbb{Z}^+ \right\}, \qquad (3.5)$$

for which no invariant complementary subspace exists. Thus the representation (3.4) on V is indecomposable. The representation subduced on every V is also indecomposable.

It is easy to see that there exists an invariant subspace chain of the space V:

$$V \equiv V^{[0]} \supset V^{[1]} \supset V^{[2]} \supset \cdots \supset V^{[N]} \supset \cdots. \qquad (3.6)$$

Correspondingly, there are quotient spaces $V^{[N,K]} = V^{[N]}/V^{[N+K]}$:

$$V^{[N,K]}: \left\{ Y(r_i,s_i) = X(r_i,s_i) \bmod V^{[N+K]} \mid N \leqslant \sum_{i=1}^{m-1} s_i \leqslant N + K - 1 \right\}, \quad N \in \mathbb{Z}^+, \quad K \in \mathbb{N}. \qquad (3.7)$$

The representation on $V^{[N]}$ can induce a representation $\bar{\Gamma}$ on $V^{[N,K]}$. When $K \geqslant 2$, the representation $\bar{\Gamma}$ on $V^{[N,K]}$ is indecomposable. When $K = 1$, the representation $\bar{\Gamma}$ on $V^{[N,1]}$ becomes

$$\overline{\Gamma}(T_p)Y(r_i,s_i)=\left(n-\sum_{k=1}^{m-1}r_k\right)\sum_{j=1}^{m-1}D(T_p)_{jm}Y(r_i+\delta_{ij},s_i)+\left(n-\sum_{k=1}^{m-1}r_k\right)D(T_p)_{mm}Y(r_i,s_i)$$
$$+\sum_{j=1}^{m-1}\sum_{k=1}^{m-1}D(T_p)_{jk}r_kY(r_i-\delta_{ik}+\delta_{ij},s_i)+\sum_{k=1}^{m-1}D(T_p)_{mk}r_kY(r_i-\delta_{ik},s_i)\qquad\left(\sum_{i=1}^{m-1}s_i=N\right).\tag{3.8}$$

It is noted that the values s_i $(i=1,2,\ldots,m-1)$ do not change. Thus every set $(s_1,s_2,\ldots,s_{m-1})$ that satisfies the condition $\Sigma_{i=1}^{m-1}s_i=N$ defines a $\overline{\Gamma}$-invariant subspace $V^{[N,1]}_{(s_1,s_2,\ldots,s_{m-1})}$ with basis

$$V^{[N,1]}_{(s_1,s_2,\ldots,s_{m-1})}:\left\{Y(r_i,s_i)\in V^{[N,1]}\,\middle|\,\sum_{i=1}^{m-1}s_i=N,s_1,s_2,\ldots,s_{m-1}\text{ are fixed}\right\}.\tag{3.9}$$

The representation on $V^{[N,1]}$ is completely reducible:

$$V^{[N,1]}=\sum_{\substack{(s_1,s_2,\ldots,s_{m-1})\\(\Sigma_{i=1}^{m-1}s_i=N)}}\oplus V^{[N,1]}_{(s_1,s_2,\ldots,s_{m-1})}.\tag{3.10}$$

When n is not a non-negative integer, the representation subduced on every $V^{[N,1]}_{(s_1,s_2,\ldots,s_{m-1})}$ is irreducible. When $n\in\mathbb{Z}^+$, it is obvious that there exists an invariant subspace $V^{[N,1]}_{(s_1,s_2,\ldots,s_{m-1})}(n)$ of $V^{[N,1]}_{(s_1,s_2,\ldots,s_{m-1})}$,

$$V^{[N,1]}_{(s_1,s_2,\ldots,s_{m-1})}(n):\left\{Y(r_i,s_i)\in V^{[N,1]}_{(s_1,s_2,\ldots,s_{m-1})}\,\middle|\,\sum_{i=1}^{m-1}r_i\leqslant n,\quad r_i\in\mathbb{Z}^+\right\},\tag{3.11}$$

with the dimension

$$\dim V^{[N,1]}_{(s_1,s_2,\ldots,s_{m-1})}(n)=\sum_{k=1}^{n}\frac{(m+k-2)!}{k!\,(m-2)!},\tag{3.12}$$

for which no invariant complementary subspace exists. Thus the representation subduced on $V^{[N,1]}_{(s_1,s_2,\ldots,s_{m-1})}$ is indecomposable. The subspace $V^{[N,1]}_{(s_1,s_2,\ldots,s_{m-1})}(n)$ carries a finite-dimensional irreducible representation.

The relation $\{a_i-\Lambda_i|\Lambda_i\in\mathbb{C},i=1,\ldots,m-1\}$ generates a left ideal J' of V. For the quotient space $\mathscr{F}'=V/J'$, also called the Fock space, a basis can be chosen as

$$\mathscr{F}':\{X(r_i)\equiv X(r_i,0)\bmod J'|r_i\in\mathbb{Z}^+\}.\tag{3.13}$$

The representation (3.4) on V induces a representation on $\mathscr{F}'$,

$$\Gamma(T_p)X(r_i)=\sum_{j=1}^{m-1}\left[\left(n-\sum_{i=1}^{m-1}r_i\right)D(T_p)_{jm}+\sum_{k=1}^{m-1}D(T_p)_{jk}\Lambda_k-D(T_p)_{mm}\Lambda_j\right]X(r_i+\delta_{ij})$$
$$-\sum_{k=1}^{m-1}\sum_{j=1}^{m-1}D(T_p)_{jm}\Lambda_kX(r_i+\delta_{ik}+\delta_{ij})+\sum_{j=1}^{m-1}\sum_{k=1}^{m-1}D(T_p)_{jk}r_kX(r_i-\delta_{ik}+\delta_{ij})$$
$$+\sum_{k=1}^{m-1}D(T_p)_{mk}r_kX(r_i-\delta_{ik})+\left[\left(n-\sum_{i=1}^{m-1}r_i\right)D(T_p)_{mm}+\sum_{k=1}^{m-1}D(T_p)_{mk}\Lambda_k\right]X(r_i),\tag{3.14}$$

which is an infinite-dimensional irreducible representation for the case with $\Lambda_i\neq0$. When $\Lambda_1=\Lambda_2=\cdots=\Lambda_{m-1}=0$, Eq. (3.14) becomes

$$\Gamma(T_p)X(r_i)=\left(n-\sum_{i=1}^{m-1}r_i\right)\sum_{j=1}^{m-1}D(T_p)_{jm}X(r_i+\delta_{ij})+\sum_{j=1}^{m-1}\sum_{k=1}^{m-1}D(T_p)_{jk}r_kX(r_i-\delta_{ik}+\delta_{ij})$$
$$+\sum_{k=1}^{m-1}D(T_p)_{mk}r_kX(r_i-\delta_{ik})+D(T_p)_{mm}\left(n-\sum_{k=1}^{m-1}r_k\right)X(r_i).\tag{3.15}$$

The representation (3.15) is equivalent to the representation on every $V^{[N,1]}_{(s_1,s_2,\ldots,s_{m-1})}$. When $n\notin\mathbb{Z}^+$, (3.15) is an infinite-dimensional representation. When $n\in\mathbb{Z}^+$, (3.15) is an indecomposable representation in which there exists a finite-dimensional irreducible representation on the invariant subspace of $\mathscr{F}'$,

$$\mathscr{F}'(n):\left\{X(r_i)\in\mathscr{F}'\,\middle|\,\sum_{i=1}^{m-1}r_i\leqslant n,\ r_i\in\mathbb{Z}^+\right\},\tag{3.16}$$

with dimension

$$\dim\mathscr{F}'(n)=\sum_{k=0}^{n}\frac{(m+k-2)!}{k!\,(m-2)!}.\tag{3.17}$$

From the above discussion we see that we can obtain the finite-dimensional representations on the subspace of Fock space if we use the IHBR of Lie algebra. If we adopt the homogeneous boson realization of Lie algebra, we can only obtain the finite-dimensional representations on the quotient spaces of Fock space (see Refs. 5 and 6).

IV. IHBR OF INDECOMPOSABLE REPRESENTATIONS OF SU(2) ALGEBRA

The representation of the one-state Heisenberg–Weyl algebra $\mathscr{H}$: $\{a^+,a,E\}$ on its universal enveloping algebra Ω with PBW basis

$$\Omega: \{X(r,s,t) \equiv a^{+r}a^sE^t \,|\, r,s,t \in \mathbb{Z}^+\} \tag{4.1}$$

is defined as

$$\begin{aligned}\rho(a^+)X(r,s,t) &= X(r+1,s,t),\\ \rho(a)X(r,s,t) &= X(r,s+1,t) + rX(r-1,s,t+1),\\ \rho(E)X(r,s,t) &= X(r,s,t+1).\end{aligned} \tag{4.2}$$

The relation $(E-1)$ generates a left ideal I of Ω. The representation (4.2) induces on the quotient space $V=\Omega/I$ with basis

$$V: \{X(r,s) \equiv X(r,s,0) \bmod I \,|\, r,s \in \mathbb{Z}^+\} \tag{4.3}$$

a representation

$$\begin{aligned}\rho(a^+)X(r,s) &= X(r+1,s),\\ \rho(a)X(r,s) &= X(r,s+1) + rX(r-1,s),\\ \rho(E) &= 1.\end{aligned} \tag{4.4}$$

By making use of the IHBR (2.25b) of su(2) algebra and the equation

$$\Gamma(T^+) = n\rho(a^+) - [\rho(a^+)]^2\rho(a), \quad \Gamma(T^-) = \rho(a), \quad \Gamma(T^0) = -n/2 + \rho(a^+)\rho(a), \tag{4.5}$$

we obtain the representation of su(2) algebra on V,

$$\begin{aligned}\Gamma(T^+)X(r,s) &= (n-r)X(r+1,s) - X(r+2,s+1),\\ \Gamma(T^-)X(r,s) &= X(r,s+1) + rX(r-1,s),\\ \Gamma(T^0)X(r,s) &= (r-n/2)X(r,s) + X(r+1,s+1).\end{aligned} \tag{4.6}$$

It is easy to see that a non-negative integer S defines an invariant subspace $V^{[S]}$ of V,

$$V^{[S]}: \{X(r,s) \,|\, s \geqslant S, \; r,s \in \mathbb{Z}^+\}, \tag{4.7}$$

for which no invariant complementary subspace exists. Thus the representation (4.6) is indecomposable. The representation subduced on every $V^{[S]}$ is also indecomposable.

On the invariant subspace chain of V,

$$V \equiv V^{[0]} \supset V^{[1]} \supset V^{[2]} \supset \cdots \supset V^{[S]} \supset \cdots, \tag{4.8}$$

there exist some quotient spaces $V^{[S,K]} = V^{[S]}/V^{[S+K]}$:

$$V^{[S,K]}: \{Y(r,s) \equiv X(r,s) \bmod V^{[S+K]} \,|\, S \leqslant s \leqslant S+K-1, \; r,s \in \mathbb{Z}^+\}, \quad S \in \mathbb{Z}^+, \quad K \in \mathbb{N}. \tag{4.9}$$

When $K \geqslant 2$, the representation induced on $V^{[S,K]}$ is indecomposable. When $K=1$, the representation induced on $V^{[S,1]}$ is

$$\begin{aligned}\Gamma(T^+)Y(r,S) &= (n-r)Y(r+1,S),\\ \Gamma(T^-)Y(r,S) &= rY(r-1,S),\\ \Gamma(T^0)Y(r,S) &= (r-n/2)Y(r,S).\end{aligned} \tag{4.10}$$

Equations (4.10) constitute an infinite-dimensional irreducible representation for the case with $n \notin \mathbb{Z}^+$. If $n \in \mathbb{Z}^+$, there exists an $(n+1)$-dimensional subspace $V^{[S,1]}(n)$ of $V^{[S,1]}$ with basis

$$V^{[S,1]}: \{Y(r,S) \in V^{[S,1]} \,|\, r \leqslant n, \; r \in \mathbb{Z}^+\} \tag{4.11}$$

for which no invariant complementary subspace exists. Thus Eqs. (4.10) are an indecomposable representation for the case with $n \in \mathbb{Z}^+$. If we define a new basis for $V^{[S,1]}(n)$,

$$|j,m\rangle_S = (1/\sqrt{(j+m)!\,(j-m)!})Y(j+m,S), \tag{4.12}$$

where $j=n/2$, $m=-j,-j+1,\ldots,j$, the representation subduced on $V^{[S,1]}(n)$ becomes

$$\begin{aligned}\Gamma(T^\pm)|j,m\rangle_S &= \sqrt{(j\mp m)(j\pm m+1)}\,|j,m\pm 1\rangle_S,\\ \Gamma(T^0)|j,m\rangle_S &= m|j,m\rangle_S,\end{aligned} \tag{4.13}$$

which is an irreducible representation of su(2) of the highest weight j with dimension $(2j+1)$.

The relation $\{a-\Lambda \,|\, \Lambda \in \mathbb{C}\}$ generates a left ideal J'. The representation (4.6) induces on the Fock space $\mathscr{F}' \equiv V/J'$,

$$\mathscr{F}': \{X(r) \equiv X(r,0) \bmod J' \,|\, r \in \mathbb{Z}^+\}, \tag{4.14}$$

a representation

$$\begin{aligned}\Gamma(T^+)X(r) &= (n-r)X(r+1) - \Lambda X(r+2),\\ \Gamma(T^-)X(r) &= \Lambda X(r) + rX(r-1),\\ \Gamma(T^0)X(r) &= (r-n/2)X(r) + \Lambda X(r+1).\end{aligned} \tag{4.15}$$

Equations (4.15) are an infinite-dimensional irreducible representation for the case with $\Lambda \neq 0$. If $\Lambda = 0$, (4.15) become

$$\begin{aligned}\Gamma(T^+)X(r) &= (n-r)X(r+1),\\ \Gamma(T^-)X(r) &= rX(r-1),\\ \Gamma(T^0)X(r) &= (r-n/2)X(r),\end{aligned} \tag{4.16}$$

which is equivalent to the representation on $V^{[S,1]}$. Equation (4.16) is an infinite-dimensional irreducible representation for the case with $n \notin \mathbb{Z}^+$. When $n \in \mathbb{Z}^+$, there exists an invariant subspace $\mathscr{F}'(n)$ of $\mathscr{F}'$ with dimension $(n+1)$,

$$\mathscr{F}'(n): \{X(r) \in \mathscr{F}' \,|\, r \leqslant n, \; r \in \mathbb{Z}^+\}, \tag{4.17}$$

for which no invariant complementary subspace exists. Thus the representation (4.16) is indecomposable for the case with $n \in \mathbb{Z}^+$. If we define a new basis for $\mathscr{F}'(n)$,

$$|j,m\rangle = (1/\sqrt{(j+m)!\,(j-m)!})X(j+m), \tag{4.18}$$

where $j=n/2$, $m=-j, -j+1,\ldots, j$, the representation subduced on $\mathscr{F}'(n)$ becomes

$$\begin{aligned}\Gamma(T^\pm)|j,m\rangle &= \sqrt{(j\mp m)(j\pm m+1)}\,|j,m\pm 1\rangle,\\ \Gamma(T^0)|j,m\rangle &= m|j,m\rangle,\end{aligned}$$

which is an irreducible representation of su(2) of the highest weight j with dimension $(2j+1)$.

For example, when $j=\frac{1}{2}$ and the order of the basis for $\mathscr{F}'$ is chosen as

$$\{|\tfrac{1}{2},\tfrac{1}{2}\rangle,|\tfrac{1}{2},-\tfrac{1}{2}\rangle,X(2),X(3),X(4),...\}, \tag{4.19}$$

in which $|\frac{1}{2},\frac{1}{2}\rangle$ and $|\frac{1}{2},-\frac{1}{2}\rangle$ are the basis for $\mathscr{F}'(n)$, the representation on $\mathscr{F}'$ $(\Lambda=0)$ is

$$\Gamma(T^+)=\left(\begin{array}{cc|cccc} 0 & 1 & & 0 & & \\ 0 & 0 & & & & \\ \hline & & 0 & 0 & 0 & \cdots \\ 0 & & -1 & 0 & 0 & \cdots \\ & & 0 & -2 & 0 & \cdots \\ & & \vdots & \vdots & \vdots & \ddots \end{array}\right),$$

$$\Gamma(T^-)=\left(\begin{array}{cc|cccc} 0 & 0 & 2 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & \cdots \\ \hline & & 0 & 3 & 0 & \cdots \\ 0 & & 0 & 0 & 4 & \cdots \\ & & 0 & 0 & 0 & \cdots \\ & & \vdots & \vdots & \vdots & \ddots \end{array}\right),$$

$$\Gamma(T^0)=\left(\begin{array}{cc|cccc} \frac{1}{2} & 0 & & 0 & & \\ 0 & -\frac{1}{2} & & & & \\ \hline & & \frac{3}{2} & 0 & 0 & \cdots \\ 0 & & 0 & \frac{5}{2} & 0 & \cdots \\ & & 0 & 0 & \frac{7}{2} & \cdots \\ & & \vdots & \vdots & \vdots & \ddots \end{array}\right),$$

which is an infinite-dimensional indecomposable representation.

ACKNOWLEDGMENT

The authors are much obliged to Professor Zhao-Yan Wu, Jilin University, People's Republic of China.

[1]L. Hlavaty and J. Niederle. Czech, J. Phys. B **29**, 283 (1979).
[2]P. A. M. Dirac, Int. J. Theor. Phys. **23**, 677 (1984).
[3]B. Gruber and R. Lenczewshi, J. Phys. A: Math. Gen. **16**, 3703 (1983).
[4]R. Lenczewshi and B. Gruber, J. Phys. A: Math. Gen. **19**, 1 (1986).
[5]Chang-Pu Sun, J. Phys. A: Math. Gen. **20**, 4551 (1987).
[6]Chang-Pu Sun, J. Phys. A: Math. Gen. **20**, 5823 (1987).
[7]Chang-Pu Sun, J. Phys. A: Math. Gen. **20**, L1157 (1987).
[8]A. V. Turbiner and A. G. Ushiveridze, Phys. Lett. A **126**, 181 (1987).
[9]A. V. Turbiner, Commun. Math. Phys. **118**, 467 (1988).
[10]M. A. Shifman and A. V. Turbiner, Commun. Math. Phys. (in press).
[11]M. A. Shifman, "New findings in quantum mechanics" preprint CERN-TH. 5265/88.

Classical non-Abelian Berry's phase

Yihren Wu
Department of Mathematics, Hofstra University, Hempstead, New York 11550

(Received 19 April 1989; accepted for publication 4 October 1989)

The classical non-Abelian Berry's phase is defined for a parameter-dependent dynamical system that is collective with respect to a Hamiltonian G-action when the parameters are fixed. It is shown that the corresponding angular two-form of Berry is given in terms of the momentum mapping of the G-action and is induced naturally by the symplectic form on the phase space. Moreover, this angular two-form can be seen as the correction term in the effective symplectic form on the parameter space, with respect to which the Hamiltonian is to be quantized.

I. INTRODUCTION

The quantum adiabatic phase, or the Berry phase,[1] is a phase shift in the eigenfunctions to a multi-parameter-dependent Hamiltonian as the parameters traverse adiabatically along a closed curve. Simon[2] has recognized that this phase arises from a connection on the solution line bundle over the parameter space. The adiabatic assumption was subsequently removed by the work of Aharonov and Anandan.[3] Applications of Berry phase arise as fractional statistics of Wilczek,[4] which include statistics for the quantum Hall effect,[5] and vortex quantization in an incompressible ^{4}He superfluid,[6] among others.

The classical Berry phase, as formulated by Berry[7] (see also Berry and Hannay[8] for the nonadiabatic generalization), is applicable for completely integrable dynamical systems. It arises as the angular shifts on the invariant tori as the system is being transported along a closed curve in the parameter space. The specific formula depends explicitly on the action variables. Following the bundle interpretation of the quantum Berry phase of Simon, one can consider the classical Berry phase as a connection form on the torus bundle over the parameter space. From the integrability assumption, this bundle is a principal $G = [SO(2)]^n$ bundle, where the G orbits are Lagrangian submanifolds on the phase space, which correspond to a one-dimensional eigensubspace via geometric quantization according to the dictionary[9] of Marsden and Weinstein. This phase is then analogous to an *Abelian* gauge field on the parameter space. In the subsequent work of Gozzi *et al.*,[10] a dynamical meaning of this phase is given. If one considers the curvature form (the angular two-form of Berry[7]) on the parameter space as a *symplectic form*, one can treat the parameters dynamically; in particular, it can be quantized. Moreover, if the parameter space is itself a symplectic manifold (the parameters being the "slow" dynamical variables), then the correct symplectic form for its quantization is the "effective form" given by the curvature as the adjustment term. This view has been applied to the planar three-body problem to yield the correct quantization.[11]

It was Wilczek and Zee[12] who first discussed Berry phase on multidimensional eigensubspaces in the quantum case, as a generalization of the work of Simon. They considered a vector bundle over the parameter space as a unitary bundle. Kristsis[13] has classified the Berry phase according to the topological type of this bundle using homotopy theory. Mention must be made, however, that the unitary group action is not a dynamical symmetry of the system, but the symmetry of a complex vector space. For a system with dynamical symmetry group G, Anandan[14] has analyzed the vector bundle for a Hamiltonian which is *collective* with respect to G and *invariant* with respect to an Abelian subgroup K (see Sec. III).

The purpose of this paper is to give the non-Abelian version of the classical Berry phase, in the presence of a dynamical symmetry group G, where the phase space M possesses a Hamiltonian G action, with G-equivariant momentum mapping $\mathbf{J}$. Our objective is to define the effective symplectic form on the parameter space which is to be used in the quantization of *G-collective* Hamiltonian functions. We also discuss the relation between the adjustment term in this effective form and the curvature form as a result of a horizontal lift. Recently, Golin *et al.*[15] have studied the situation for *G-invariant* Hamiltonians, where the momentum mapping is defined with respect to the effective form. We summarize our work as follows: in Sec. II we reformulate the standard (Abelian) classical Berry's angular two-form in the terminology of symplectic geometry to facilitate the non-Abelian generalization. We also explain the interpretation of Berry's angular two-form as a symplectic form, as a kind of reduction procedure. It is with this point of view that we extend our consideration to the non-Abelian case. The needed concepts in symplectic geometry: momentum mapping and collective motion, are collected in Sec. III. The non-Abelian Berry phase is formulated in Sec. IV. For completeness, we include proofs of some technical details in the Appendix.

II. ABELIAN BERRY PHASE-SYMPLECTIC REFORMULATION

We take the viewpoint that Berry phase is an adjustment term in a separation of variables type scheme. Denote by X the total phase space which separates into $X = M \times B$, where M and B are symplectic manifolds with symplectic forms Ω_M and Ω_B, respectively. We also consider B as a parameter space of a Hamiltonian

$$H_b(m) = H(m,b): X \to \mathbb{R}. \tag{1}$$

We assume in this section that for each $b \in B$, the dynamical

 0022-2488/90/020294-06$03.00

system defined by $H_b:M\to\mathbb{R}$ is integrable, i.e., there exist action-angle variables $(I_{i,b},\theta_{i,b})$, $i=1,\ldots,n=\frac{1}{2}\dim M$, with

$$\dot{I}_{i,b}=0,\quad \dot{\theta}_{i,b}=\frac{\partial H_b}{\partial I_{i,b}}. \tag{2}$$

Let $\mathbf{J}:M\times B\to\mathbb{R}^n$ be defined by $\mathbf{J}(m,b)=(I_1,\ldots,I_n)$, where we consider $I_i(m,b)=I_{i,b}(m)$ as functions on $M\times B$. Let $E=\mathbf{J}^{-1}(\mu)$ for some $\mu\in\mathbb{R}^n$. We have natural projections:

$$\begin{array}{ccc} & E & \\ \pi\swarrow & & \searrow\rho, \\ B & & M \end{array}$$

where $\pi:E\to B$ is a principal $G=[\mathrm{SO}(2)]^n$ bundle (by the angle variables). Moreover, fibers of π are Lagrangian submanifolds in M. Therefore, $\Omega_M=0$ when restricted to $\pi^{-1}(b)$. However, in general, $\Omega_M\neq 0$ when restricted to E (it is zero if the system is separable, I_i independent of B, for instance).

We use the convention that the group G acts on the left. The projection π gives rise to a diffeomorphism $\tilde{\pi}:G\setminus E\to B$, it is *not* a canonical transformation. Since the Hamiltonian vector field of H commutes with the G action, it defines a vector field on the quotient space $G\setminus E$. The correct symplectic form for this dynamical system is the one induced by $\Omega_M+\Omega_B$ on $G\setminus E$. However, since the differential form Ω_M (a contravariant object) does not push forward onto the quotient space, we perform an average over the G orbits. The extent to which this method works depends on whether the method of average[16] yields reasonable results. This includes, but is not limited to, adiabatic systems. We assume that the resulting two-form is nondegenerate, therefore it is a symplectic form. Pull back via the diffeomorphism $\tilde{\pi}^{-1}$ gives a two-form on B;

$$\Omega_B^{(\mathrm{eff})}=\Omega_B+\Omega_A, \tag{3}$$

where Ω_A carries the effect of Ω_M on $G\setminus E$. Then $\tilde{\pi}$ is a canonical transform if we change the symplectic form on B by adding Ω_A, where Ω_A is the average of Ω_M over the fibers, which coincides with the angular two-form of Berry. Explicitly,

$$\begin{aligned}\Omega_A(b)&=\pi_*\rho^*\Omega_M(m,b)\\ &=(2\pi)^{-n}\int_G g^*\rho^*\Omega_M(g(m,b))dg.\end{aligned} \tag{4}$$

Here ρ^* denotes the pull back operator and π_* denotes the averaging operator over the fiber $\pi^{-1}(b)$. This average is well defined since $g^*\Omega(g(x))$ is a two-form at x for all $g\in G$, two-forms at x forms a vector space. This effective form also plays a role in the covariant constant condition in geometric quantization.[11]

According to Berry, the angular phase shift along a closed curve γ in B is given by

$$\Delta\theta_i=\frac{-\partial}{\partial I_i}\int_D\Omega_A, \tag{5}$$

where D is a region with boundary γ. We may view

$$\Omega_\nabla=-\sum\frac{\partial}{\partial I_i}\Omega_A e_i \tag{6}$$

as the curvature form of a connection α_∇ on the principal bundle E, where $\{e_i\}$ forms a basis for the Lie algebra of $[\mathrm{SO}(2)]^n=\mathbb{R}^n$. It is then a tautology that $\Delta\theta_i$ is the phase shift of the horizontal lift of γ with respect to α_∇. Thus we recover the principal bundle interpretation of Simon.[2]

The significance of the horizontal lift, in the context of symplectic geometry, can be seen as follows: Since the motion on B along a vector field $\mathscr{V}$ will affect the change in the angular variables, Berry[7] considered the effective Hamiltonian function

$$\widetilde{H}=H-\int_G p\mathscr{V}q\,d\theta=H-\mathscr{V}\lrcorner\,\pi_*\rho^*\alpha_M, \tag{7}$$

where α_M is the symplectic one-form on M, then $\dot{\theta}=\partial\widetilde{H}/\partial I$. The second term, as a result of the horizontal lift of $\mathscr{V}$, accounts for the change in θ caused by $\mathscr{V}$. In a sense, this is changing the Hamiltonian vector field $[(\partial\widetilde{H}/\partial I)(\partial/\partial\theta)$ instead of $(\partial H/\partial I)(\partial/\partial\theta)]$ of the function H. This in turn changes the Poisson bracket and the symplectic structure on M. If we are only concerned with the dynamics on M (the dynamics on B is given by the curve γ), then it is natural to define adjustments on M. In our work, we leave M as is and put the adjustments on B, where more interesting dynamics may occur.[11] In summary, we see that given a Hamiltonian $H:M\times B\to\mathbb{R}$, on fixing the actions $\mathbf{J}=\mu$, we have $H:\mathbf{J}^{-1}(\mu)\to\mathbb{R}$, which induces a Hamiltonian $F:G\setminus\mathbf{J}^{-1}(\mu)=B\to\mathbb{R}$. To quantize $F:B\to\mathbb{R}$, we must use the corrected symplectic form $\Omega_B^{(\mathrm{eff})}$ on B.

III. MOMENTUM MAPPING AND COLLECTIVE MOTION

We give here a brief review of some concepts of symplectic geometry that are needed for our purpose, details may be found in Guillemin and Sternberg.[17]

Let M be a symplectic manifold with symplectic form $\Omega_M=d\alpha_M$. Here, M has a compact Lie group G acting canonically on the left, i.e., $g^*\Omega_M=\Omega_M$. Denote by $\mathbf{g}$ its Lie algebra and we assume that the first two cohomologies of $\mathbf{g}$ vanish; this is the case if G is semi-simple. Here, M is a Hamiltonian G space if there exist a *momentum mapping* $\mathbf{J}:M\to\mathbf{g}^*$, the dual space of $\mathbf{g}$, which satisfies:

(i) $\mathbf{J}$ is G equivariant, i.e., $\mathbf{J}(gm)=\mathrm{ad}^*_{g^{-1}}\mathbf{J}(m)$, where $\mathrm{ad}^*_{g^{-1}}$ denotes the coadjoint action of g on $\mathbf{g}^*$.

(ii) Let $\xi\in\mathbf{g}$, by the canonical G action on M and ξ induces a Hamiltonian vector field $\xi^{\#}$ on M. We require that $\xi^{\#}$ is the Hamiltonian vector field of the real value function $f(m)=\langle\mathbf{J}(m),\xi\rangle$, where $\langle,\rangle$ is the pairing between $\mathbf{g}$ and $\mathbf{g}^*$. If we denote by $\mathscr{H}_f$ the Hamiltonian vector field of f, then $\mathscr{H}_f(m)=\xi^{\#}(m)$.

Note that in the integrable case, the actions $\mathbf{J}:M\to[\mathbf{so}^*(2)]^n\cong\mathbb{R}^n$, satisfies the above conditions.

Let $H:M\to\mathbb{R}$ be a Hamiltonian function, H is *collective* with respect to G if there exists a map $F:\mathbf{g}^*\to\mathbb{R}$ such that $H(m)=F(\mathbf{J}(m))$. This corresponds to the quantum Hamiltonian operator belonging to the Lie algebra. In particular, if G is Abelian, H is G collective and this implies that H is G invariant.

Contrary to the integrable case, $\mathbf{J}^{-1}(\mu)$ has no quantum analog, e.g., if $G=\mathrm{SO}(3)$, $M=T^*\mathbb{R}^3$, $\mathbf{J}^{-1}(\mu)$ consists of a subset with fixed angular momentum, this violates the uncertainty principle. Fix μ_0 in $\mathbf{g}^*$, let S be the coadjoint orbit of μ_0 in $\mathbf{g}^*$, where S has a natural symplectic structure Ω_S. We assume that $\mathbf{J}^{-1}(S)$ is a coisotropic submanifold in M. (This assumption usually holds.) Reduction on coisotropic submanifold is the classical analog of projection onto an eigensubspace according to Marsden and Weinstein.[9] In the case with Hamiltonian G action, this assumption corresponds to restriction to a $\mathbf{g}$-representation subspace. In the example above, this is the eigensubspace with the magnitude of the angular momentum fixed. The reduction procedure goes as follows: Let K be the connected component of the isotropy subgroup of μ_0, then H, when restricted to $\mathbf{J}^{-1}(S)$, is invariant with respect to K in the following sense: For any $m \in \mathbf{J}^{-1}(S)$, $\mathbf{J}(m) = \mathrm{ad}^*_{g^{-1}}(\mu_0)$ for some $g \in G$, then for all $k \in K$, $H(gkg^{-1}m) = H(m)$. This does not constitute a new K action on M due to the choice of g, however, K orbits of this "action" is well defined. Denote the space of K orbits by $K \setminus \mathbf{J}^{-1}(S)$, this space has a natural symplectic form Ω satisfying $p^*\Omega = i^*\Omega_M$, where p is the projection $\mathbf{J}^{-1}(S) \to K \setminus \mathbf{J}^{-1}(S)$ and i is the inclusion $\mathbf{J}^{-1}(S) \to M$. The Hamiltonian vector field of H on $\mathbf{J}^{-1}(S)$ is related to the Hamiltonian vector field of F on S by $\mathbf{J}_* \mathscr{H}_H(m) = \mathscr{H}_F(\mathbf{J}(m))$. Since the Hamiltonian vector fields are related, the quantization of H on $K \setminus \mathbf{J}^{-1}(S)$ is determined by the quantization of F on $S = G/K$, which is usually of much lower dimension.

The manifold $K \setminus \mathbf{J}^{-1}(\mu_0)$ is also a symplectic manifold with symplectic form Ω_0 satisfying $p^*\Omega_0 = i^*\Omega_M$, where $p:\mathbf{J}^{-1}(\mu_0) \to K \setminus \mathbf{J}^{-1}(\mu_0)$ and $i:\mathbf{J}^{-1}(\mu_0) \to M$ are projection and inclusion, respectively.

Locally we may decompose

$$T_m M = [T_m \mathbf{J}^{-1}(\mu_0)]^\perp \oplus T_m K \setminus \mathbf{J}^{-1}(\mu_0), \tag{8a}$$

where $^\perp$ denotes symplectic orthogonal complement. This decomposition is not canonical in a categorical sense. We emphasize here that although K is not assumed to be Abelian, K actions Poisson commute on $\mathbf{J}^{-1}(\mu_0)$, i.e., $[\mathbf{k},\mathbf{k}] \subseteq \ker \mu_0$, thus

$$[T_m \mathbf{J}^{-1}(\mu_0)]^\perp = \mathbf{g} \oplus \mathbf{k}^* = \mathbf{k} \oplus \mathbf{s} \oplus \mathbf{k}^*, \tag{8b}$$

where $\mathbf{s}$ is the orthogonal complement of $\mathbf{k}$ in $\mathbf{g}$, to emphasize that it is locally S, and we reserve $^\perp$ to mean symplectic orthogonal complement. Here, $\mathbf{k}^*$ can be interpreted either as the dual coordinates of $\mathbf{k}$ in the symplectic manifold T^*K, or as the dual Lie algebra of $\mathbf{k}$. Since $\mathbf{s}$ forms a symplectic vector space under $\mu_0[-,-]$, we may assume the basis e_i's forms a canonical coordinate system on $\mathbf{s}$, i.e., let $\{e_i\}$ $i = 1,\ldots,2n$ be a basis for $\mathbf{s}$, $\{e_{2n+i}\}$ basis for $\mathbf{k}$, we have

$$\begin{aligned} &\mu_0[e_{n+i},e_j] = \delta_{i,j}, \\ &\mu_0[e_i,e_j] = \mu_0[e_{n+i},e_{n+j}] = 0, \quad \text{for } i,j = 1,\ldots,n, \\ &\mu_0[e_{2n+i},e_j] = 0, \quad \text{for all } i,j. \end{aligned} \tag{9}$$

Therefore, although $\mathbf{J}e_i$'s are not canonical coordinates on M, the $\mathbf{s}$ components have the canonical relation at $m \in \mathbf{J}^{-1}(\mu_0)$. Moreover, since K actions $\mathbf{J}e_{2n+i}$ are in involution, there exist angle variables θ_{2n+i} for these actions. Thus locally, M is a twisted produce of three symplectic manifolds

$$M = T^*K \times S \times K / \mathbf{J}^{-1}(\mu_0). \tag{10}$$

We indicate a method, mod $\mathbf{k}$, for such a decomposition:

Lemma: Let $\mathbf{J}(m) = \mu_0$ and K be the isotropy subgroup of μ_0 as usual, let $f:M \to \mathbb{R}$ be a K-invariant (not necessarily G-collective) function around m. Let

$$\mathscr{H}_f = (\mathscr{H}_f - \overline{\mathscr{H}}_f) + \overline{\mathscr{H}}_f,$$

where

$$\overline{\mathscr{H}}_f = \mathscr{H}_f - \sum_i \mathscr{H}_{\mathbf{J}_{n+i}}(f)\mathscr{H}_{\mathbf{J}_i} - \mathscr{H}_{\mathbf{J}_i}(f)\mathscr{H}_{\mathbf{J}_{n+i}} \tag{11}$$

is the desired decomposition, i.e., $\overline{\mathscr{H}}_f \in T_m \mathbf{J}^{-1}(\mu_0)$ is K equivariant, thus it projects via p_* to $T_m K \setminus \mathbf{J}^{-1}(\mu_0)$ and $\mathscr{H}_f - \overline{\mathscr{H}}_f \in \mathbf{s}$. We will henceforth compress the notation $\mathbf{J}e_i = \mathbf{J}_i$. We are now in position to explain (6), the relation between the effective symplectic form and the curvature form. As was mentioned in Ref. 11, our work is inspired by a result of Kummer,[18] which we now describe: Let X be a manifold with K acting freely on the left, let $Y = K \setminus X$ be the manifold of orbits. Let α be a connection on the principal bundle $\pi:X \to Y$. We can extend the K action into a Hamiltonian K action on T^*X. Denote by $\mathbf{J}:T^*X \to \mathbf{k}^*$ the momentum mapping, $\mathbf{J}$ is linear. We assume, as in our case, μ_0 is K invariant. Then

$$\tilde{\pi}:K \setminus \mathbf{J}^{-1}(\mu_0) \to T^*Y \tag{12}$$

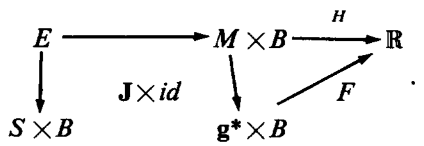
is a diffeomorphism. (This plays the role of $\tilde{\pi}:G \setminus E \to B$ in the integrable case.) Moreover, $\tilde{\pi}$ is a canonical transform if we have the effective symplectic form $\Omega_{T^*Y} + \mu_0\Omega_\alpha$. Here Ω_α is the curvature of the connection α; it is a $\mathbf{k}$ valued two-form. $\mu_0 \in \mathbf{k}^*$ implies $\mu_0\Omega_\alpha$ is a real valued two-form on T^*Y. Thus $\mu_0\Omega_\alpha$ plays the role of Ω_A, the adjustment term. In this sense, relation (6) is the inverse relation of

$$\Omega_A = \mu_0\Omega_\alpha = \mathbf{J}\Omega_\alpha. \tag{6'}$$

In the Abelian case, (6) and (6′) are equivalent. In our case, since $\Omega_\alpha = d\alpha + \frac{1}{2}[\alpha,\alpha]$, and $\mu_0[\alpha,\alpha] = 0$, relation (6) will only recover part of Ω_α.

IV. NON-ABELIAN BERRY PHASE

We assume M,B are symplectic manifolds with symplectic forms $\Omega_M = d\alpha_M$ and Ω_B, respectively, as before. Here and beyond, we denote $\mathscr{H}_\phi$ the Hamiltonian vector field of the function ϕ_b on M, viewed as a vector field on $M \times B$. Suppose for each $b \in B$, there is a Hamiltonian G action on M with momentum mapping $\mathbf{J}_b:M \to \mathbf{g}^*$, denote $\mathbf{J}:M \times B \to \mathbf{g}^*$ as usual. We assume $H(m,b):M \times B \to \mathbb{R}$ be such that H_b is G collective. That is, for each $b \in B$, we have $F_b:\mathbf{g}^* \to \mathbb{R}$ with $H_b(m) = F_B(\mathbf{J}_b(m))$. More conveniently, the following diagram commutes:



Let S be the coadjoint orbit of μ_0 as before, $E = \mathbf{J}^{-1}(S)$, K be the isotropy subgroup of μ_0, then H, when restricted to E, is determined by $F:S \times B \to \mathbb{R}$, where $S \times B$ has a natural symplectic form $\Omega_S + \Omega_B$. However, as seen in the Abelian case (S is a point), this symplectic form has to be adjusted in order to quantize F.

Since G-collective Hamiltonians are in involution with the K action, these actions are adiabatic invariants. So we assume here that the vector fields $\mathscr{V}$ on B we will consider satisfies an "adiabatic" condition $\mathscr{V}(\mathbf{J}_{2n+i}) \sim O(\epsilon)$, ϵ small. This will hold either if the evolution on B is slow, or the dependence of $\mathbf{J}$ on B is small. We are seeking Hamiltonian equations which will describe:

$$\dot{\mathbf{J}}_{2n+i} = 0, \quad \dot{\mathbf{J}}_i = \text{level crossing},$$

$$\dot{\theta}_{2n+i} = \text{phase shift in } \mathbf{k} \text{ representation subspace}.$$

Our assumption on K implies the irreducible representation spaces of K are one dimensional. Denote by $K\backslash E$ the space of K orbits, define

$$W: K\backslash E \rightarrow S \times K\backslash \mathbf{J}^{-1}(\mu_0), \tag{13}$$

and

$$W(K(m,b)) = (\mathbf{J}(m,b) = \mathrm{ad}^*_{g^{-1}}(\mu_0), Kg^{-1}(m,b))$$

is a diffeomorphism. Furthermore, denote by $\Omega_{K\backslash E}$ the induced form of $\Omega_M + \Omega_B$ on $K\backslash E$ via restriction and average, it pulls back to a form which separates into three terms.

Theorem 1:

$$\Omega_{K\backslash E}(K(m,b)) = W^*[\Omega_S(\mathrm{ad}^*_{g^{-1}}(\mu_0)) + \Omega_{\mathrm{cross}} + \Omega_0(g^{-1}(m,b))], \tag{14}$$

where Ω_0 is a two-form on $K\backslash \mathbf{J}^{-1}(\mu_0)$ and Ω_S is the natural symplectic form on S. Here Ω_{cross} is a pull back from $S\times B$, $\Omega_{\mathrm{cross}}(\tilde{e}_i, v) = v(\mathbf{J}e_i)$, $\Omega_{\mathrm{cross}} = 0$ when restricted to either factor. Moreover, $p^*\Omega_0 = p_* i^*(\Omega_M + \Omega_B)$ with $p: \mathbf{J}^{-1}(\mu_0) \rightarrow K\backslash \mathbf{J}^{-1}(\mu_0)$ and $i: \mathbf{J}^{-1}(\mu_0) \rightarrow M \times B$ natural projection and inclusion, p_* the average operator over K orbits. Finally, $p_* i^* \Omega_B$ can be identified naturally with $i^*\Omega_B$.

With the above separation, the adjustment term Ω_A on $S\times B$ should capture the effect of $p_* i^* \Omega_M$ (the only term not described in the theorem) on $\mathbf{J}^{-1}(\mu_0)$. The effective symplectic form on $S\times B$ is then $\Omega^{\mathrm{eff}}_{S\times B} = \Omega_s + \Omega_B + \Omega_A + \Omega_{\mathrm{cross}}$ with respect to which F is quantized. The quantum effect of the symplectic form Ω_S on S has been studied by Giavarini and Onofri[19] in terms of coherent states and reproducing kernel. This will handle level crossing due to evolution on S. (In Ref. 19, where S is the parameter space, there is no B.) Clearly, since the two-form on S will have no effect on $\mathbf{J}^{-1}(\mu_0)$, this implies Ω_A is a two-form on B. Unlike the integrable case, $p_* i^* \Omega_M$ does not "push forward" to a two-form on B. This is due to the fact that the projection $\tilde{\pi}: K\backslash \mathbf{J}^{-1}(\mu_0) \rightarrow B$ has nontrivial fibers, whereas in the integrable case $\tilde{\pi}$ is a diffeomorphism. However, the lemma gives a way to subtract off the fibers $K\backslash \mathbf{J}_b^{-1}(\mu_0)$. Therefore the required adjustment on B must capture the effect of Ω_M on $[T_m \mathbf{J}_b^{-1}(\mu_0)]^\perp$, the symplectic orthogonal complement of $T_m \mathbf{J}_b^{-1}(\mu_0)$ in $T_m M$. We identify $T_m \mathbf{J}_b^{-1}(\mu_0) = T_{(m,b)} \mathbf{J}_b^{-1}(\mu_0)$, where $\mathbf{J}_b^{-1}(\mu_0)$ is viewed as a submanifold in $\mathbf{J}^{-1}(\mu_0) \rightarrow M \times B$.

The symplectic manifold M is a twisted product of three symplectic manifolds each of whose symplectic structure is induced by Ω_M. We have local decomposition, as suggested by (10), around m, where $\mathbf{J}(m,b) = \mu_0$:

$$M = T^*K \times S \times K\backslash \mathbf{J}_b^{-1}(\mu_0). \tag{10'}$$

All three symplectic forms depend on B, and we will measure the change of the symplectic forms on $T^*K \times S$ along a vector field on B, the last factor $K\backslash J_b^{-1}(\mu_0)$ being irrelevant in the realm of collective Hamiltonians. Theorem 1 allows us to restrict our discussion to the submanifold $\mathbf{J}^{-1}(\mu_0)$, on which the momentum $\mathbf{J}_i$ $(i = 1,...,2n)$ are canonical. Roughly,

$$\Omega_M(m,b) = d_M \mathbf{J}_{n+i} \wedge d_M \mathbf{J}_i + d_M \mathbf{J}_{2n+i} \wedge d_M \boldsymbol{\theta}_{2n+i},$$

so

$$-\Omega_A = d_B \mathbf{J}_{n+i} \wedge d_B \mathbf{J}_i + d_B \mathbf{J}_{2n+i} \wedge d_B \boldsymbol{\theta}_{2n+i}. \tag{15}$$

We will discuss the relation between this adjustment form and the curvature form for the lift. This is done by first lifting vector fields from B to the relevant piece in $K\backslash \mathbf{J}^{-1}(\mu_0)$, then by lifting to $\mathbf{J}^{-1}(\mu_0)$. This is, along the vertical diagram,

$$\begin{array}{ccc} \mathbf{J}^{-1}(\mu_0) & \xrightarrow{\rho} & M \\ \downarrow p & & \\ K\backslash \mathbf{J}^{-1}(\mu_0) & & \\ \downarrow \pi & & \\ B & & \end{array}$$

We first consider the lift $K\backslash \mathbf{J}^{-1}(\mu_0) \rightarrow B$, denoted by

$$\begin{aligned} &\mathscr{R}(K(m,b)) \\ &\quad = p_* \rho_*^{-1}(T_m \mathbf{J}_b^{-1}(\mu_0))^\perp \\ &\quad = \{v \in T_{(m,b)} K\backslash \mathbf{J}^{-1}(\mu_0) \mid \text{ there exists} \\ &\qquad \tilde{v} \in T_{(m,b)} \mathbf{J}^{-1}(\mu_0), \quad p_* \tilde{v} = v, \\ &\qquad \text{and } \Omega_M(\rho_* \tilde{v}, \tilde{w}) = 0, \\ &\qquad \text{for all } \tilde{w} \in T_{(m,b)} \mathbf{J}_b^{-1}(\mu_0)\}, \end{aligned} \tag{16}$$

the horizontal subspace in $T_{(m,b)} K\backslash \mathbf{J}^{-1}(\mu_0)$. It is clear that under the projection $\pi: K\backslash \mathbf{J}^{-1}(\mu_0) \rightarrow B, \pi_* \mathscr{R}(K(m,b)) \rightarrow T_b B$ is one-to-one and onto, since $\mathscr{R}(K(m,b)) \cap T_{(m,b)} K\backslash \mathbf{J}_b^{-1}(\mu_0) = 0$ by the nondegeneracy of the symplectic form. Thus $\mathscr{R}$ defines a horizontal space for lifting from B to $K\backslash \mathbf{J}^{-1}(\mu_0)$. This isomorphism π_* plays the role of the diffeomorphism $\tilde{\pi}: G\backslash E \rightarrow B$ in the integrable case, and the effect of Ω_M on $\mathscr{R}$ induces a two-form on B.

Note that if α_∇ is the connection form for the lift from B to $K\backslash \mathbf{J}^{-1}(\mu_0)$, and β_∇ is the connection form for the lift from $K\backslash \mathbf{J}^{-1}(\mu_0)$ to $\mathbf{J}^{-1}(\mu_0)$, then $p^*\alpha_\nabla + \beta_\nabla$ is the connection form for the lift from B to $\mathbf{J}^{-1}(\mu_0)$. Thus the combined curvature form, as well as the "angular two-form," also sum.

We first state the results for the $K\backslash \mathbf{J}^{-1}(\mu_0)$ component. Here

$$\alpha_\nabla(K(m,b)): T_{K(m,b)} K\backslash \mathbf{J}^{-1}(\mu_0) \rightarrow T_{K(m,b)} K\backslash \mathbf{J}_b^{-1}(\mu_0), \tag{17}$$

such that $\alpha_\nabla =$ identity on $T_{K(m,b)} K\backslash \mathbf{J}_b^{-1}(\mu_0)$, and $\alpha_\nabla = 0$ on $\mathscr{R}(K(m,b))$. Let v be vector fields on B and satisfying the adiabatic assumption, $\tilde{v}$ be K equivariant on $\mathbf{J}^{-1}(\mu_0)$

in the preimage of $\mathscr{R}$ so that its projection via p_* is the horizontal lift of v. Then, mod **k**

$$\tilde{v} = v + \sum_i v(\mathbf{J}_{n+i})\mathscr{H}_{\mathbf{J}_i} - v(\mathbf{J}_i)\mathscr{H}_{\mathbf{J}_{n+i}}. \quad (18)$$

Theorem 2: Let $\{e_i\}_{i=1,\ldots,n,\ldots 2n}$ be a basis for **s** in **g** which forms a canonical coordinate for **s** as a symplectic vector space, then

$$\Omega_A = \sum_i (d_B \mathbf{J}_i) \wedge (d_B \mathbf{J}_{n+i}) \quad (19)$$

is the two-form induced by Ω_M on B through $\mathscr{R}$ on $K/\mathbf{J}^{-1}(\mu_0)$. Moreover, the curvature form Ω_∇ for the connection $\mathscr{R}$ is

$$\Omega_\nabla(v,w) = \mathscr{H}_f - \sum_i \mathscr{H}_{\mathbf{J}_{n+i}}(f)\mathscr{H}_{\mathbf{J}_i} - \mathscr{H}_{\mathbf{J}_i}(f)\mathscr{H}_{\mathbf{J}_{n+i}}, \quad (20)$$

where

$$f = \sum_i v\mathbf{J}_{n+i} w\mathbf{J}_i - w\mathbf{J}_{n+i} v\mathbf{J}_i = \Omega_A(v,w).$$

Compare with the lemma, $\Omega_\nabla(v,w)$ is the vertical component of $\mathscr{H}_f$, whereas in the integrable case, set $\mathbf{J}_{n+i} = p_i$, $\mathbf{J}_i = q_i$, $\Omega_\nabla(v,w)$ is indeed $\mathscr{H}_f$, where $f =$ average over torus of

$$-\sum_i d_B p_i d_B q_i(v,w) = \Omega_A(v,w),$$

$\mathscr{H}_f$ is already vertical.

As for the K component, let $\mathscr{R}'(m,b) = p_*^{-1} \times \mathscr{R}(K(m,b))$ define the connection form $\beta_\nabla(m,b): \mathscr{R}'(m,b) \to \mathbf{k} =$ tangent space of K orbit at (m,b) by

$$\beta_\nabla(\tilde{w}) = -\sum_i \tilde{w}(\theta_{2n+i})\mathscr{H}_{\mathbf{J}_{2n+i}}, \quad (21)$$

and denote the horizontal space $\mathscr{R}^\#(m\cdot b) = \ker \beta_\nabla$, thus the horizontal lift $v^\#$ of v on B takes the form

$$v^\# = \tilde{v} + \sum_i \tilde{v}(\theta_{2n+i})\mathscr{H}_{\mathbf{J}_{2n+i}}. \quad (22)$$

Here, $\tilde{v}$ as in (18) and $v^\#$ may be viewed as the horizontal projection of $\tilde{v}$.

Notice that although θ_{2n+i} are not uniquely defined, $\tilde{v}(\theta_{2n+i})$ are when restricted to $\mathbf{J}^{-1}(\mu_0)$. However, since we do not have the angular variables at all, we will give an alternate definition of β which does not involve θ.

$$\beta_\nabla(\tilde{w}) = \sum_i \frac{\partial}{\partial \mathbf{J}_{2n+i}}(\tilde{w} \lrcorner\, p_*\rho^*\alpha_M)\mathscr{H}_{\mathbf{J}_{2n+i}}, \quad (23)$$

where $\tilde{w} \in \mathscr{R}'$, $p_*\rho^*$ pulls α_M back to the submanifold $\mathbf{J}^{-1}(\mu_0)$ and average over the K orbits. This definition is comparable to $(\partial/\partial I_k)(p\mathscr{V}q)(\partial/\partial\theta_k)$ in the integrable case. Equation (21) implies directly that $\beta_\nabla(\mathscr{H}_{\mathbf{J}_{2n+i}}) = \mathscr{H}_{\mathbf{J}_{2n+i}}$, thus it is indeed a connection form. The curvature form is given by the standard formula:

$$\tilde{\Omega}_\nabla = d\beta_\nabla + \tfrac{1}{2}[\beta_\nabla, \beta_\nabla], \quad (24)$$

as a **k** valued two-form on $K \backslash \mathbf{J}^{-1}(\mu_0)$. The following theorem is immediate:

Theorem 3: Let $v^\#, w^\#$ be horizontal projections of $\tilde{v}$ and $\tilde{w}$, respectively, define the effective form on $K \backslash \mathbf{J}^{-1}(\mu_0)$ by

$$\tilde{\Omega}_A(\tilde{v},\tilde{w}) = -p_*\rho^*\Omega_M(v^\#,w^\#). \quad (25)$$

Then

$$d\beta_\nabla(\tilde{v},\tilde{w}) = \mathscr{H}_f, \quad (26)$$

where $f = \tilde{\Omega}_A(\tilde{v},\tilde{w})$. Notice that the commutator term in the curvature form is not recovered as explained at the end of Sec. III.

To summarize: Eq. (25) is precisely statement (15) that the effective form on B is the one induced by Ω_M. The horizontal lift of v on B is given by (22) and (18), the curvature form is partially recovered from the effective form by (26) and (20).

APPENDIX: PROOFS OF THEOREMS 1–3

Proof of Theorem 1: By equivariance of momentum mapping, we have the Poisson bracket on M,

$$\{\mathbf{J}\xi, \mathbf{J}\eta\} = \mathbf{J}[\xi,\eta] = \mu_0[\xi,\eta], \quad (\text{A1})$$

for all ξ,η in **g**. The first term is $\Omega_M(\mathscr{H}_{\mathbf{J}\xi}, \mathscr{H}_{\mathbf{J}\eta})$, the last term is $\Omega_S(\mathbf{J}_{b*}\mathscr{H}_{\mathbf{J}\xi}, \mathbf{J}_{b*}\mathscr{H}_{\mathbf{J}\eta})$. It remains to compute the cross term: Let $v = v_M + v_B$ be a K-equivariant vector field on $\mathbf{J}^{-1}(\mu_0)$ and $w = \mathbf{J}_{b*}\mathscr{H}_{\mathbf{J}\xi}$ be a vector field on S, then $\Omega_M(\mathscr{H}_{\mathbf{J}\xi}, v_M) = -v_M(\mathbf{J}\xi) = v_B(\mathbf{J}\xi)$. Thus the cross term Ω_{cross} is a pull be of $S \times B$ since it depends only on the B component of v.

Proof of Theorem 2: Using (18), we have

$$\tilde{v} = v + \sum_i v(\mathbf{J}_{n+i})\mathscr{H}_{\mathbf{J}_i} - v(\mathbf{J}_i)\mathscr{H}_{\mathbf{J}_{n+i}},$$
$$\tilde{w} = w + \sum_i w(\mathbf{J}_{n+i})\mathscr{H}_{\mathbf{J}_i} - w(\mathbf{J}_i)\mathscr{H}_{\mathbf{J}_{n+i}}, \quad (18')$$

where the second terms are the M components of $\tilde{v}$ and $\tilde{w}$. Since the $\mathbf{J}_i$ are canonical, $\Omega_M(\mathscr{H}_{\mathbf{J}_{n+i}}, \mathscr{H}_{\mathbf{J}_i}) = \delta_{i,j}$, a straightforward calculation shows $\Omega_M(\rho_*\tilde{v}, \rho_*\tilde{w}) = \Sigma_i v(\mathbf{J}_i)w(\mathbf{J}_{n+i}) - v(\mathbf{J}_{n+i})w(\mathbf{J}_i)$. Moreover, $\Omega_M(\mathbf{g},\mathbf{k}) = 0$ implies it is independent of our choice of $\tilde{v}$ and $\tilde{w}$, proving (19).

Since $\Omega_\nabla(v,w) = \alpha_\nabla[\tilde{v},\tilde{w}]$, we compute

$$[\tilde{v},\tilde{w}] = [v,w] + \sum_i [v,w]\mathbf{J}_{n+i}\mathscr{H}_{\mathbf{J}_i} - [v,w]\mathbf{J}_i\mathscr{H}_{\mathbf{J}_{n+i}}$$
$$+ \mathscr{H}_f - \sum_i \mathscr{H}_{\mathbf{J}_{n+i}}(f)\mathscr{H}_{\mathbf{J}_i} - \mathscr{H}_{\mathbf{J}_i}(f)\mathscr{H}_{\mathbf{J}_{n+i}}$$
$$+ \sum v(\mathbf{J}_i)w(\mathbf{J}_j) - v(\mathbf{J}_j)w(\mathbf{J}_i)[\mathscr{H}_{\mathbf{J}_k}, \mathscr{H}_{\mathbf{J}_l}]. \quad (\text{A2})$$

Here f as in the theorem and α_∇ of the first line is 0 since it is horizontal. The adiabatic assumption implies f is K invariant, thus α_∇ of the second line is itself since it is vertical. As for the last line, notice that $[\mathscr{H}_{\mathbf{J}_k}, \mathscr{H}_{\mathbf{J}_l}]$ is on $\mathbf{J}_b^{-1}(\mu_0)$, and

$$[\mathscr{H}_{\mathbf{J}_k}, \mathscr{H}_{\mathbf{J}_l}] = \mathscr{H}_{\mathbf{J}\{e_k e_l\}}, \quad (\text{A3})$$

where $\mathbf{J}[e_k,e_l] = \mu_0[e_k,e_l] =$ constant on $\mathbf{J}^{-1}(\mu_0)$. Therefore

$$\Omega_M(\mathscr{H}_{\mathbf{J}[e_k,e_l]},\xi)=\xi(\mathbf{J}[e_k,e_l])=0, \qquad \text{(A4)}$$

for all $\xi\in T_m\mathbf{J}_b^{-1}(\mu_0)$. This implies $p_*\mathscr{H}_{\mathbf{J}[e_k,e_l]}\in\mathscr{R}$, $\alpha_\nabla[\mathscr{H}_{\mathbf{J}_k},\mathscr{H}_{\mathbf{J}_l}]=0$.

Proof of Theorem 3: Equations (25) and (26) are straightforward once we have established (23) as our working definition of β_∇. We must therefore show (23) is equivalent to (21), and this can be seen as

$$p_*\rho^*\alpha_M=\sum_i\mathbf{J}_{n+i}d_M\mathbf{J}_i+\sum_k\mathbf{J}_{2n+k}d_M\theta_{2n+k}$$

if we have the angular variables. Thus

$$\sum_k\mathbf{J}_{2n+k}\widetilde{w}(\theta_{2n+k})=\widetilde{w}\lrcorner\Big(p_*\rho^*\alpha_M-\sum_i\mathbf{J}_{n+i}d_M\mathbf{J}_i\Big),$$

the last expression does not involve θ's. One checks that $\widetilde{w}\lrcorner\mathbf{J}_{n+i}d_M\mathbf{J}_i=0$ using the canonical relations $\mathbf{J}_i$'s (9), and the result follows.

[1] M. V. Berry, Proc. R. Soc. London, Ser. A **392**, 45 (1984).
[2] B. Simon, Phys. Rev. Lett. **51**, 2167 (1983).
[3] Y. Aharonov and J. Anandan, Phys. Rev. Lett. **58**, 1593 (1987).
[4] F. Wilczek, Phys. Rev. Lett. **49**, 957 (1982).
[5] D. Arovas, J. R. Schrieffer, and F. Wilczek, Phys. Rev. Lett. **53**, 722 (1984).
[6] F. D. M. Haldane and Y. Wu, Phys. Rev. Lett. **55**, 2887 (1985).
[7] M. V. Berry, J. Phys. A: Math. Gen. **18**, 15 (1985).
[8] M. V. Berry and J. H. Hannay, J. Phys. A: Math. Gen. **21**, L325 (1988).
[9] J. Marsden and A. Weinstein, Bull. Am. Math. Soc. **1**, 545 (1979).
[10] E. Gozzi and W. Thacker, Phys. Rev. D **35**, 2398 (1987).
[11] Y. Wu, J. Phys. A: Math. Gen. **22**, L117 (1989).
[12] F. Wilczek and A. Zee, Phys. Rev. Lett. **52**, 2111 (1984).
[13] E. Kristsis, Commun. Math. Phys. **111**, 417 (1987).
[14] J. Anandan, Phys. Lett. A **129**, 201 (1988).
[15] S. Golin, A. Knauf, and S. Marmi, "The Hannay angles: geometry, adiabaticity, and an example," preprint (1988).
[16] V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978).
[17] V. W. Guillemin and S. Sternberg, *Symplectic Techniques in Physics* (Cambridge U.P., Cambridge, 1984).
[18] M. Kummer, Indiana U. Math. J. **30**, 281 (1981).
[19] G. Giavarini and E. Onofri, J. Math. Phys. **30**, 659 (1989).

The group theoretical analysis of gravitational instanton equations

M. Przanowski
Institute of Physics, Technical University of Łódź, Wólczańska 219, 93-005 Łódź, Poland

(Received 3 May 1989; accepted for publication 2 August 1989)

Group theoretical properties of certain nonlinear partial differential equations playing a distinguished role in the gravitational instanton theory and in complex relativity are studied. It is demonstrated that, in general, the groups of contact transformations admitted by these equations appear to be the first prolongations of appropriate point transformation groups. An exceptional case leading to the Gibbons–Hawking metric is examined in detail.

I. INTRODUCTION

The group theoretical analysis of nonlinear partial differential equations provides us with an effective tool for finding solutions of these equations.[1–6] First, having the maximal group of point or contact transformations leaving the differential equation invariant one can generate new solutions of this equation from some given solution. It may happen that all solutions can be obtained in this manner (equivalent to the automorphic equation). For instance, it occurs when the differential equation appears to be linearizable by the point or contact transformation.[6] Second, employing the method of invariant variables,[2–6] one reduces the number of independent variables. In particular, the application of this method in general relativity leads to space-times with the Killing vector field.[6]

Consequently, motivated by the above facts we intend to analyze the point and contact symmetries of certain nonlinear partial differential equations that play an important role in both the gravitational instanton theory and complex relativity. These equations arise from the reduction of ten Einstein equations $R_{ij} = 0$ for the spaces of Petrov–Penrose–Plebański types $[-]\otimes[\text{any}]$, $D\otimes[\text{any}]$, or of the equations $R_{ij} = -\Lambda g_{ij}$ for the spaces $[-]\otimes[\text{any}]$.[7–15] For each case one gets a single second-order, nonlinear partial differential equation for one function.

The group theoretical analysis of the equations obtained shows that, except the Gibbons–Hawking case,[16–20] the maximal groups of contact transformations admitted by these equations appear to be the first prolongations of the point transformation groups. Then the transformation of metric caused by the appropriate maximal group of contact transformations is rather trivial (except the Gibbons–Hawking metric) as it consists of a simple coordinate transformation and the conformal transformation of a constant conformal factor (compare Ibragimov[6,21] and Pham Mau Quan[22]). In the case of the Gibbons–Hawking metric the corresponding nonlinear equation is linearizable. Therefore every Gibbons–Hawking metric can be obtained from one seed metric with the use of the appropriate contact transformation group.

The paper is organized as follows: Section II is devoted to the heavenly equation and its counterpart in the gravitational instanton theory. In this section we also present the formalism of Lie–Bäcklund transformations in the space of infinite-order jets,[5,6,23] which is employed in our paper. In Sec. III, the nonzero cosmological constant is included. In Sec. IV, we deal with one-sided type-D, Ricci-flat complex space-times and gravitational instantons. Then we consider a nonlinear partial differential equation [Eq. (4.12)] that contains the Gibbons–Hawking case and the case of heavens, called "case III,"[18] admitting the Killing vector field. In the gravitational instanton theory "case III" corresponds to the gravitational instantons admitting the "rotational" Killing vector field.[19,20,24]

In the next paper we analyze the group theoretical properties of the fundamental equations of complex relativity, i.e., the hyperheavenly equations.

II. HEAVENS AND SELF-DUAL, RICCI-FLAT GRAVITATIONAL INSTANTONS

In the present section we deal with the "first heavenly equation" and its counterpart in the gravitational instanton theory.[7–13]

The first heavenly equation is one of the fundamental partial differential equations in complex relativity. It is of the form

$$\frac{\partial^2 u}{\partial x^1\,\partial x^3}\frac{\partial^2 u}{\partial x^2\,\partial x^4} - \frac{\partial^2 u}{\partial x^1\,\partial x^4}\frac{\partial^2 u}{\partial x^2\,\partial x^3} - 1 = 0, \quad (2.1)$$

where $u = u(x^i)$, $i = 1,\dots,4$, is a holomorphic function of four complex variables x^1, x^2, x^3, and x^4. It is well known that every self-dual, Ricci-flat complex space-time $\mathscr{H}$ (for heaven) admits local complex coordinates x^1, x^2, x^3, and x^4 such that the metric g of $\mathscr{H}$ is given locally as follows:

$$g = g_{\alpha\bar\beta}\cdot(dx^\alpha\otimes dx^{\bar\beta} + dx^{\bar\beta}\otimes dx^\alpha), \quad (2.2a)$$

$$g_{\alpha\bar\beta} = \frac{\partial^2 u}{\partial x^\alpha\,\partial x^{\bar\beta}}, \quad (2.2b)$$

with $\alpha = 1,2$, and $\bar\beta = 3,4$, and where $u = u(x^1)$ is a solution of (2.1).

Analogously, every self-dual, Ricci-flat gravitational instanton admits local complex coordinates x^1, x^2, $x^3 = \bar x^1$, $x^4 = \bar x^2$ (with the overbar standing for the complex conjugation) such that the metric of this gravitational instanton can be defined locally by (2.2) with $u = u(x^i)$ now being a real solution of Eq. (2.1).

Consequently the group theoretical properties of the first heavenly equation (2.1) determine immediately the symmetries of the gravitational instanton version of (2.1). Therefore we intend to find point and contact groups of transformations under which Eq. (2.1) remains invariant and then we will specify the results obtained to the case of gravitational instanton theory.

 0022-2488/90/020300-08$03.00

As announced in the Introduction, we employ the infinite jet bundle technique.[5,6,23] Using the notation

$$u_{i_1\cdots i_s} := \frac{\partial^s u}{\partial x^{i_1}\cdots\partial x^{i_s}}, \quad s\geqslant 1, \quad i_1,\ldots,i_s = 1,2,3,4, \quad (2.3)$$

we rewrite (2.1) in the form

$$F := u_{13}u_{24} - u_{14}u_{23} - 1 = 0. \quad (2.4)$$

Equation (2.4) defines a submanifold $\mathscr{F}$ of the space of second-order jets of $C^4 \to C, J^2(C^4,C)$. The infinite prolongation of (2.4) defines a submanifold $\mathscr{F}^\infty$ of the space of infinite-order jets, $J^\infty(C^4,C)$. Let G_1 be a one-parameter Lie–Bäcklund group of transformations of $J^\infty(C^4,C)$ and let X be the infinitesimal operator of G_1:

$$X = \xi^i \frac{\partial}{\partial x^i} + \eta \frac{\partial}{\partial u} + \sum_{s\geqslant 1} \zeta_{i_1\cdots i_s} \frac{\partial}{\partial u_{i_1\cdots i_s}},$$

$$\xi^i = \xi^i(x^j,u,u_j,u_{j_1j_2},\ldots), \quad \eta = \eta(x^j,u,u_j,u_{j_1j_2},\ldots), \quad (2.5)$$

$$\zeta_{i_1\cdots i_s} = D_{i_1}\cdots D_{i_s}(\eta - \xi^j u_j) + \xi^j u_{ji_1\cdots i_s},$$

where

$$D_i := \frac{\partial}{\partial x^i} + u_i \frac{\partial}{\partial u} + \sum_{s\geqslant 1} u_{ii_1\cdots i_s} \frac{\partial}{\partial u_{i_1\cdots i_s}} \quad (2.6)$$

(indices $i,j,i_1,\ldots,i_s,j_1,j_2,\ldots$ are assumed to run through 1,2,3,4). Then the differential equation (2.4) is invariant under the group G_1 iff

$$XF\,|_{\mathscr{F}^\infty} = 0, \quad (2.7)$$

where $|_{\mathscr{F}^\infty}$ means the restriction to the submanifold $\mathscr{F}^\infty$. An operator of the form (2.5) is called a Lie–Bäcklund operator. Two Lie–Bäcklund operators X_1 and X_2 are said to be equivalent if there exist functions $\rho^i = \rho^i(x^j,u,u_j,u_{j_1j_2},\ldots)$ such that

$$X_1 - X_2 = \rho^i D_i. \quad (2.8)$$

Hence one concludes that the Lie–Bäcklund operator (2.5) is equivalent to

$$X_c := \mu \frac{\partial}{\partial u} + \sum_{s\geqslant 1} D_{i_1}\cdots D_{i_s}(\mu) \frac{\partial}{\partial u_{i_1\cdots i_s}}, \quad (2.9)$$

where

$$\mu = \eta - \xi^i u_i. \quad (2.10)$$

This X_c is called a canonical Lie–Bäcklund operator equivalent to X.

Now as any partial differential equation is invariant under the one-parameter group defined by the infinitesimal operator of the form $\rho^i D_i$, Eq. (2.7) holds iff

$$X_c F\,|_{\mathscr{F}^\infty} = 0. \quad (2.11)$$

In the present paper we deal with groups of contact transformations. Consequently, we study Eq. (2.11) for X_c of the form (2.9) with μ now being a function of the variables x^i, u, and u_i, i.e., $\mu = \mu(x^i,u,u_i)$. Then the infinite prolongation of the infinitesimal operator of a one-parameter group of contact transformations leaving Eq. (2.4) invariant is defined by (2.5) with

$$\xi^i = -\frac{\partial\mu}{\partial u_i}, \quad \eta = \mu - u_i \frac{\partial\mu}{\partial u_i} \quad (2.12)$$

(compare Ref. 6). It is evident that this group of contact transformations appears to be the first prolongation of some one-parameter group of point transformations iff

$$\frac{\partial^2\mu}{\partial u_i\,\partial u_j} = 0. \quad (2.13)$$

One finds easily that the invariance condition (2.11) for F defined by (2.4) and X_c defined by (2.9) takes the form of

$$[u_{24}D_1D_3(\mu) + u_{13}D_2D_4(\mu) - u_{23}D_1D_4(\mu) - u_{14}D_2D_3(\mu)]_{\mathscr{F}^\infty} = 0. \quad (2.14)$$

A straightforward but rather tedious computation leads to the following general solution of (2.14):

$$\mu = \alpha(x^1,x^2) + \beta(x^3,x^4) + au + \left[\frac{\partial\gamma(x^1,x^2)}{\partial x^2} - bx^1\right]u_1 + \left[-\frac{\partial\gamma(x^1,x^2)}{\partial x^1} - bx^2\right]u_2 + \left[\frac{\partial\delta(x^3,x^4)}{\partial x^4} - (a-b)x^3\right]u_3 + \left[-\frac{\partial\delta(x^3,x^4)}{\partial x^3} - (a-b)x^4\right]u_4, \quad (2.15)$$

where $\alpha(x^1,x^2)$, $\beta(x^3,x^4)$, $\gamma(x^1,x^2)$, and $\delta(x^3,x^4)$ are arbitrary holomorphic functions of their arguments, and a and b are any complex constants.

As μ given by (2.15) satisfies the condition (2.13), every group of contact transformations admitted by Eq. (2.4) is the first prolongation of the point transformation group admitted by (2.4).

According to (2.12) with (2.15),

$$\begin{aligned}
\xi^1 &= -\frac{\partial\gamma(x^1,x^2)}{\partial x^2} + bx^1, \quad \xi^2 = \frac{\partial\gamma(x^1,x^2)}{\partial x^1} + bx^2,\\
\xi^3 &= -\frac{\partial\delta(x^3,x^4)}{\partial x^4} + (a-b)x^3,\\
\xi^4 &= \frac{\partial\delta(x^3,x^4)}{\partial x^3} + (a-b)x^4,\\
\eta &= \alpha(x^1,x^2) + \beta(x^3,x^4) + au.
\end{aligned} \quad (2.16)$$

Having ξ^i and η one can integrate the Lie equations and find the maximal group of contact transformations leaving Eq. (2.4) invariant. Thus we arrive at the following theorem.

Theorem 2.1: The maximal group of contact transformations admitted by Eq. (2.4) is the first prolongation of the infinite group of point transformations defined by

$$\begin{aligned}
x^{1'} &= \omega^{1'}(x^1,x^2), \quad x^{2'} = \omega^{2'}(x^1,x^2),\\
x^{3'} &= \omega^{3'}(x^3,x^4), \quad x^{4'} = \omega^{4'}(x^3,x^4),\\
u' &= cu + \sigma(x^1,x^2) + \tau(x^3,x^4),
\end{aligned} \quad (2.17)$$

where $\sigma(x^1,x^2)$ and $\tau(x^3,x^4)$ are arbitrary holomorphic functions of their arguments, c is an arbitrary nonzero complex constant, and the $\omega^{i'}$ are any holomorphic functions of their arguments satisfying the following condition:

$$\det\left|\left|\frac{\partial\omega^{\alpha'}}{\partial x^{\alpha}}\right|\right|\cdot\det\left|\left|\frac{\partial\omega^{\bar\beta'}}{\partial x^{\bar\beta}}\right|\right| = c^2,$$
$$\alpha = 1,2,\quad \bar\beta = 3,4. \qquad \blacksquare$$

The point transformation (2.17) causes the following transformation of the heavenly metric (2.2):

$$g_{\alpha\bar\beta}\mapsto g'_{\alpha'\bar\beta'} = c g_{\alpha\bar\beta}\frac{\partial x^{\alpha}}{\partial x^{\alpha'}}\ \frac{\partial x^{\bar\beta}}{\partial x^{\bar\beta'}}. \tag{2.18}$$

Consequently we conclude that (i) one cannot generate essentially new heavenly metrics from some original metric with the use of contact transformations; and (ii) one cannot linearize Eq. (2.4) by any contact transformation (compare Ref. 6).

These conclusions seem to clarify the difficulties in searching for solutions of the first heavenly equation (2.4).

Now it is an easy matter to specify the results obtained to the case of a self-dual, Ricci-flat gravitational instanton; namely, in the latter case Theorem 2.1 remains valid with the obvious restrictions:

$$\begin{aligned} &x^3 = \overline{x^1},\quad x^4 = \overline{x^2},\quad \bar u = u,\\ &\omega^{3'} = \overline{\omega^{1'}},\quad \omega^{4'} = \overline{\omega^{2'}},\quad \bar c = c,\quad \bar\sigma = \tau, \end{aligned} \tag{2.19}$$

where an overbar stands for complex conjugation.

III. HEAVENS WITH COSMOLOGICAL CONSTANT

We intend to generalize the results of Sec. II when the nonzero cosmological constant is included. As has been shown in Ref. 14, the Einstein equations $R_{ij} = -\Lambda g_{ij}$, $\Lambda\neq 0$, for a self-dual complex (or Euclidean) Einstein space-time can be brought locally to one nonlinear partial differential equation on one holomorphic (resp. real) function u:

$$F := u_{13}u_{24} - u_{14}u_{23} - (2u_{13} + u_1u_3)e^{-u} = 0. \tag{3.1}$$

Then the metric is defined by (2.2a) with

$$\begin{aligned} &g_{13} = 3\Lambda^{-1}u_{13},\quad g_{14} = 3\Lambda^{-1}u_{14},\\ &g_{23} = 3\Lambda^{-1}u_{23},\quad g_{24} = 3\Lambda^{-1}\cdot(u_{24} - 2e^{-u}). \end{aligned} \tag{3.2}$$

To find the group of contact transformations admitted by Eq. (3.1) we proceed as in Sec. II; namely, we solve Eq. (2.11) for X_c and F given by (2.9) and (3.1), respectively, and for $\mu = \mu(x^i,u,u_i)$. One finds that the invariant condition (2.11) is now of the form

$$\begin{aligned} [&(u_{24} - 2e^{-u})D_1D_3(\mu) + u_{13}D_2D_4(\mu) - u_{23}D_1D_4(\mu)\\ &- u_{14}D_2D_3(\mu) - u_3e^{-u}D_1(\mu) - u_1e^{-u}D_3(\mu)\\ &+ \mu\cdot(2u_{13} + u_1u_3)e^{-u}]_{\mathcal{F}^\infty} = 0. \end{aligned} \tag{3.3}$$

The general solution of Eq. (3.3) reads

$$\begin{aligned} \mu = &\frac{\partial\alpha(x^2)}{\partial x^2} + \frac{\partial\beta(x^4)}{\partial x^4} + \gamma(x^1,x^2)u_1 - \alpha(x^2)u_2\\ &+ \delta(x^3,x^4)u_3 - \beta(x^4)u_4, \end{aligned} \tag{3.4}$$

where $\alpha(x^2)$, $\beta(x^4)$, $\gamma(x^1,x^2)$, and $\delta(x^3,x^4)$ are arbitrary holomorphic functions of their arguments. The function μ given by (3.4) satisfies the condition (2.13). Consequently, every group of contact transformations admitted by Eq. (3.1) appears to be the first prolongation of an appropriate group of point transformations admitted by (3.1). Then we have

$$\begin{aligned} &\xi^1 = -\gamma(x^1,x^2),\quad \xi^2 = \alpha(x^2),\\ &\xi^3 = -\delta(x^3,x^4),\quad \xi^4 = \beta(x^4),\\ &\eta = \frac{\partial\alpha(x^2)}{\partial x^2} + \frac{\partial\beta(x^4)}{\partial x^4}, \end{aligned} \tag{3.5}$$

and the integration of Lie equations yields the following theorem.

Theorem 3.1: The maximal group of contact transformations admitted by Eq. (3.1) is the first prolongation of the infinite group of point transformations admitted by (3.1):

$$\begin{aligned} &x^{1'} = \omega^{1'}(x^1,x^2),\quad x^{2'} = \omega^{2'}(x^2),\\ &x^{3'} = \omega^{3'}(x^3,x^4),\quad x^{4'} = \omega^{4'}(x^4),\\ &u' = u + \ln\left(\frac{\partial\omega^{2'}}{\partial x^2}\frac{\partial\omega^{4'}}{\partial x^4}\right), \end{aligned} \tag{3.6}$$

where the $\omega^{i'}$ are any holomorphic functions of their arguments such that

$$\frac{\partial\omega^{1'}}{\partial x^1}\frac{\partial\omega^{2'}}{\partial x^2}\frac{\partial\omega^{3'}}{\partial x^3}\frac{\partial\omega^{4'}}{\partial x^4}\neq 0. \qquad \blacksquare$$

It is obvious that statements (i) and (ii) of Sec. II also hold true in the present case. To specify the results obtained to the case of gravitational instantons one should assume that $x^3 = \overline{x^1}$, $x^4 = \overline{x^2}$, $\bar u = u$, $\omega^{3'} = \overline{\omega^{1'}}$, and $\omega^{4'} = \overline{\omega^{2'}}$.

IV. ONE-SIDED TYPE-D, RICCI-FLAT COMPLEX SPACE-TIMES AND GRAVITATIONAL INSTANTONS

The main result of Ref. 15 can be summarized as follows: For every Ricci-flat complex (or Euclidean) space-time of the type $D\otimes$[any], the Einstein equations can be reduced locally to the following nonlinear partial differential equation of one holomorphic (resp. real) function:

$$\begin{aligned} &u = u(x^1 + x^3,x^2,x^4),\\ &u_{13}u_{24} - u_{14}u_{23} - 2\cdot(u_{13} + 2u_1u_3)e^{-u} = 0. \end{aligned} \tag{4.1}$$

The metric is of the form (2.2a) with

$$\begin{aligned} &g_{13} = \tfrac{1}{2}u_1^{-3/2}u_{13},\quad g_{14} = \tfrac{1}{2}u_1^{-3/2}u_{14},\\ &g_{23} = \tfrac{1}{2}u_1^{-3/2}u_{23},\quad g_{24} = \tfrac{1}{2}u_1^{-3/2}\cdot(u_{24} - 2e^{-u}). \end{aligned} \tag{4.2}$$

We define new variables

$$y^1 := x^1 + x^3,\quad y^2 := x^2,\quad y^3 := x^4, \tag{4.3}$$

and u is now assumed to be a function of these variables $u = u(y^1,y^2,y^3)$. Consequently, Eq. (4.1) is equivalent to the following differential equation in the second-order jet space $J^2(C^3,C)$:

$$F := u_{11}u_{23} - u_{12}u_{13} - 2\cdot(u_{11} + 2u_1^2)e^{-u} = 0. \tag{4.4}$$

The formalism employed in previous sections can be, *mutatis mutandis*, utilized in the present case of $J^2(C^3,C)$. Formulas (2.5)–(2.13) remain valid if one substitutes x for y and lets the indices $i,j,i_1,\ldots,i_s,j_1,\ldots,j_s$ run through 1,2,3. The invariant condition (2.11) for Eq. (4.4) takes the form

$$\begin{aligned} [&(u_{23} - 2e^{-u})D_1D_1(\mu) + u_{11}D_2D_3(\mu) - u_{13}D_1D_2(\mu)\\ &- u_{12}D_1D_3(\mu) - 8u_1e^{-u}D_1(\mu)\\ &+ 2\mu\cdot(u_{11} + 2u_1^2)e^{-u}]_{\mathcal{F}^\infty} = 0. \end{aligned} \tag{4.5}$$

As we deal with groups of contact transformations we let μ depend on y^i, u, and u_i only. The analysis of Eq. (4.5) is, rather surprisingly, much more involved than that of Eq. (2.14) or (3.3). Finally we arrive at the following general solution of (4.5):

$$\mu = \frac{\partial \alpha(y^2)}{\partial y^2} + \frac{\partial \beta(y^3)}{\partial y^3} + (ay^1 + \gamma(y^2) + \delta(y^3))u_1 - \alpha(y^2)u_2 - \beta(y^3)u_3, \tag{4.6}$$

where $\alpha(y^2), \beta(y^3), \gamma(y^2)$, and $\delta(y^3)$ are arbitrary holomorphic functions of their arguments and a is an arbitrary complex constant.

Then

$$\begin{aligned} &\xi^1 = -ay^1 - \gamma(y^2) - \delta(y^3),\\ &\xi^2 = \alpha(y^2), \quad \xi^3 = \beta(y^3),\\ &\eta = \frac{\partial \alpha(y^2)}{\partial y^2} + \frac{\partial \beta(y^3)}{\partial y^3}. \end{aligned} \tag{4.7}$$

Having ξ^i and η one can find the maximal group of contact transformations leaving Eq. (4.4) invariant. This group is evidently the first prolongation of the maximal group of point transformations admitted by (4.4). Thus we have the following theorem.

Theorem 4.1: The maximal group of contact transformations admitted by Eq. (4.4) is the first prolongation of the infinite point transformation group

$$\begin{aligned} &y^{1'} = by^1 + \sigma(y^2) + \tau(y^3), \quad y^{2'} = \omega^{2'}(y^2),\\ &y^{3'} = \omega^{3'}(y^3),\\ &u' = u + \ln\left(\frac{\partial \omega^{2'}}{\partial y^2}\frac{\partial \omega^{3'}}{\partial y^3}\right), \end{aligned} \tag{4.8}$$

where $\sigma(y^2), \tau(y^3), \omega^{2'}(y^2)$, and $\omega^{3'}(y^3)$ are arbitrary holomorphic functions of their arguments such that

$$\frac{\partial \omega^{2'}}{\partial y^2}\frac{\partial \omega^{3'}}{\partial y^3} \neq 0,$$

and b is an arbitrary nonzero complex constant. ■

The metric (4.2) transforms under (4.8) as follows:

$$g_{\alpha\bar\beta} \mapsto g_{\alpha'\bar\beta'} = b^{3/2} g_{\alpha\bar\beta}\frac{\partial x^\alpha}{\partial x^{\alpha'}}\frac{\partial x^{\bar\beta}}{\partial x^{\bar\beta'}}, \quad \alpha = 1,2, \quad \bar\beta = 3,4, \tag{4.9}$$

where

$$\begin{aligned} &x^{1'} = bx^1 + \tfrac{1}{2}\cdot(\sigma(x^2) + \tau(x^4)), \quad x^{2'} = \omega^{2'}(x^2),\\ &x^{3'} = bx^3 + \tfrac{1}{2}\cdot(\sigma(x^2) + \tau(x^4)), \quad x^{4'} = \omega^{3'}(x^4). \end{aligned} \tag{4.10}$$

The transformation (4.9) is obviously a composition of the transformation generated by the coordinate transformation (4.10) and the conformal transformation

$$g_{\alpha\bar\beta} \mapsto b^{3/2} g_{\alpha\bar\beta}, \quad b = \text{const} \neq 0.$$

Therefore one can repeat conclusions (i) and (ii) of Sec. II as they also hold true in the present case.

It is an easy matter to specify the above-obtained results to the case of a one-sided, type-D, Ricci-flat gravitational instanton. Namely, we have to make the restrictions

$$\begin{aligned} &x^3 = \overline{x^1}, \; x^4 = \overline{x^2} \Rightarrow \overline{y^1} = y^1, \; y^3 = \overline{y^2};\\ &\bar u = u, \bar b = b, \quad \tau(y^3) = \overline{\sigma(y^2)}, \quad \omega^{3'}(y^3) = \overline{\omega^{2'}(y^2)}. \end{aligned} \tag{4.11}$$

Now we intend to study symmetries of the following nonlinear partial differential equation in $J^2(C^3,C)$:

$$F := u_{11}u_{23} - u_{12}u_{13} - f(y^1) = 0, \tag{4.12}$$

where $f(y^1)$ is an arbitrary holomorphic (or real for the instanton case) function of y^1. Equations of this type seem to play an important role in complex relativity and gravitational instanton physics.

For $f(y^1) = e^{y^1}$, Eq. (4.12) defines all heavens admitting the Killing vector field that belong to "case III" (according to the terminology of Ref. 18), or, in the gravitational instanton theory, it defines all self-dual, Ricci-flat gravitational instantons admitting the "rotational" Killing vector field.[19,20,24] As has been demonstrated in Ref. 20, Eq. (4.12) for $f(y^1) = e^{y^1}$ is related to Eq. (4.4) by some Lie–Bäcklund transformation.

For $f(y^1) = 1$, Eq. (4.12) defines all heavens admitting the Killing vector field that are classified as "case I a,"[18] or in the gravitational instanton theory it defines all self-dual, Ricci-flat gravitational instantons admitting the "translational" Killing vector field[19,20,24] (that is, the Gibbons–Hawking gravitational instantons[16,17]). Notice that in the gravitational instanton theory these two cases, $f(y^1) = e^{y^1}$ or $f(y^1) = 1$, contain all self-dual, Ricci-flat gravitational instantons admitting at least one Killing vector field.

The invariant condition (2.11) for Eq. (4.12) reads

$$[u_{23}D_1D_1(\mu) + u_{11}D_2D_3(\mu) - u_{13}D_1D_2(\mu) - u_{12}D_1D_3(\mu)]_{\mathcal{F}^*} = 0. \tag{4.13}$$

As before we let μ be a function of y^i, u, and u_i. Then the general solution of Eq. (4.13) appears to be of the form

$$\mu = \alpha(y^2,u_1,u_2) + \beta(y^3,u_1,u_3) + \gamma(y^2,y^3,u_1)u + \delta(y^i,u_1), \tag{4.14}$$

where $\alpha(y^2,u_1,u_2), \beta(y^3,u_1,u_3), \gamma(y^2,y^3,u_1)$, and $\delta(y^i,u_1)$ are any holomorphic functions satisfying the following set of equations:

$$\begin{aligned} &f\frac{\partial^2\alpha}{\partial u_2^2} = 0, \quad f\frac{\partial^2\beta}{\partial u_3^2} = 0, \quad \frac{\partial f}{\partial y^1}\frac{\partial \alpha}{\partial u_1} = 0, \quad \frac{\partial f}{\partial y^1}\frac{\partial \beta}{\partial u_1} = 0,\\ &\frac{\partial f}{\partial y^1}\frac{\partial \gamma}{\partial u_1} = 0, \quad f\frac{\partial^2\alpha}{\partial u_1^2} + \frac{\partial \gamma}{\partial y^3} = 0, \quad f\frac{\partial^2\beta}{\partial u_1^2} + \frac{\partial \gamma}{\partial y^2} = 0,\\ &\frac{\partial^2\delta}{\partial y_1^2} = 0, \quad f\frac{\partial^2\gamma}{\partial u_1^2} + \frac{\partial^2\gamma}{\partial y^2\partial y^3} = 0, \quad f\frac{\partial^2\delta}{\partial u_1^2} + \frac{\partial^2\delta}{\partial y^2\partial y^3} = 0,\\ &2f\frac{\partial \alpha}{\partial u_1} - u_1\frac{\partial \gamma}{\partial y^3} - \frac{\partial^2\delta}{\partial y^1\partial y^3} = 0,\\ &2f\frac{\partial \beta}{\partial u_1} - u_1\frac{\partial \gamma}{\partial y^2} - \frac{\partial^2\delta}{\partial y^1\partial y^2} = 0,\\ &f\cdot\left(2\gamma + 2\frac{\partial^2\delta}{\partial y^1\partial u_1} + 2u_1\frac{\partial \gamma}{\partial u_1} + \frac{\partial \alpha}{\partial y^2} + \frac{\partial \beta}{\partial y^3}\right) + \frac{\partial f}{\partial y^1}\frac{\partial \delta}{\partial u_1} = 0. \end{aligned} \tag{4.15}$$

The analysis of Eq. (4.15) leads to the results

$$\text{(A)} \quad f(y^1) = a\cdot(y^1)^b, \quad a = \text{const} \neq 0, \quad b = \text{const} \neq 0. \tag{4.16}$$

Then

$$\mu = (-c_1u_1 + c_2)y_1 + \rho(y^2) + \omega(y^3) - (c_3y^2 + c_4)u_2 - \{[2c_5 - (2+b)c_1 - c_3]y^3 + c_6\}u_3 + c_5u, \tag{4.17}$$

where $\rho(y^2)$ and $\omega(y^3)$ are arbitrary holomorphic functions of y^2 or y^3, respectively, and $c_1,...,c_6$ are arbitrary complex constants.

Hence

$$\xi^1 = c_1 y^1, \quad \xi^2 = c_3 y^2 + c_4,$$
$$\xi^3 = [2c_5 - (2+b)c_1 - c_3] + c_6, \tag{4.18}$$
$$\eta = c_5 u + c_2 y^1 + \rho(y^2) + \omega(y^3).$$

Having ξ^i and η we can find the maximal group of contact transformations admitted by Eq. (4.12) for $f(y^1)$ given by (4.16). This group is the first prolongation of the following group of point transformations:

$$y^{1\prime} = r_1 y^1, \quad y^{2\prime} = r_2 y^2 + r_3,$$
$$y^{3\prime} = r_4^2 r_1^{-(2+b)} r_2^{-1} y^3 + r_5, \tag{4.19}$$
$$u' = r_4 u + r_6 y^1 + \sigma(y^2) + \tau(y^3),$$

where $\sigma(y^2)$ and $\tau(y^3)$ are arbitrary holomorphic functions of their arguments and $r_1 \neq 0$, $r_2 \neq 0$, $r_4 \neq 0$, r_3, r_5, and r_6 are any complex constants.

We also have

$$\text{(B)} \quad f(y^1) = ae^{by^1}, \quad a = \text{const} \neq 0, \quad b = \text{const} \neq 0. \tag{4.20}$$

In this case,

$$\mu = c_1 y^1 + \rho(y^2) + \omega(y^3)$$
$$- b^{-1} \cdot \left(\frac{\partial \gamma(y^2)}{\partial y^2} + \frac{\partial \delta(y^3)}{\partial y^3} + 2c_2 \right) u_1$$
$$+ \gamma(y^2) u_2 + \delta(y^3) u_3 + c_2 u, \tag{4.21}$$

where $\rho(y^2)$, $\omega(y^3)$, $\gamma(y^2)$, and $\delta(y^3)$ are arbitrary holomorphic functions of their arguments and c_1 and c_2 are arbitrary complex constants. Consequently

$$\xi^1 = b^{-1} \cdot \left(\frac{\partial \gamma(y^2)}{\partial y^2} + \frac{\partial \delta(y^3)}{\partial y^3} + 2c_2 \right),$$
$$\xi^2 = -\gamma(y^2), \quad \xi^3 = -\delta(y^3), \tag{4.22}$$
$$\eta = c_2 u + c_1 y^1 + \rho(y^2) + \omega(y^3),$$

and one concludes that the maximal group of contact transformations admitted by Eq. (4.12) for $f(y^1)$ given by (4.20) is the first prolongation of the point transformation group

$$y^{1\prime} = y^1 + r_1 - b^{-1} \ln\left(\frac{\partial \varphi(y^2)}{\partial y^2} \frac{\partial \psi(y^3)}{\partial y^3} \right),$$
$$y^{2\prime} = \varphi(y^2), \quad y^{3\prime} = \psi(y^3), \tag{4.23}$$
$$u' = e^{1/2br_1} u + r_2 y^1 + \sigma(y^2) + \tau(y^3),$$

where $\varphi(y^2)$, $\psi(y^3)$, $\sigma(y^2)$, and $\tau(y^3)$ are arbitrary holomorphic functions of their arguments restricted only by the condition

$$\frac{\partial \varphi(y^2)}{\partial y^2} \frac{\partial \psi(y^3)}{\partial y^3} \neq 0,$$

and r_1 and r_2 are any complex constants.

Also

$$\text{(C)} \quad f(y^1) \neq 0, \quad \frac{\partial f(y^1)}{\partial y^1} \neq 0,$$

$f(y^1)$ is neither type (A) nor (B).

In this case,

$$\mu = c_1 y^1 + \rho(y^2) + \omega(y^3) - (c_2 y^2 + c_3) u_2$$
$$- [(2c_4 - c_2) y^3 + c_5] u_3 + c_4 u, \tag{4.24}$$
$$\xi^1 = 0, \quad \xi^2 = c_2 y^2 + c_3, \quad \xi^3 = (2c_4 - c_2) y^3 + c_5,$$
$$\eta = c_4 u + c_1 y^1 + \rho(y^2) + \omega(y^3), \tag{4.25}$$

where $\rho(y^2)$ and $\omega(y^3)$ are arbitrary holomorphic functions of y^2 and y^3, respectively, and $c_1,...,c_5$ are arbitrary complex constants.

Then, the maximal group of contact transformations leaving Eq. (4.12) invariant is the first prolongation of the group of point transformations

$$y^{1\prime} = y^1, \quad y^{2\prime} = r_1 y^2 + r_2, \quad y^{3\prime} = r_3^2 r_1^{-1} y^3 + r_4,$$
$$u' = r_3 u + r_5 y^1 + \sigma(y^2) + \tau(y^3), \tag{4.26}$$

where $\sigma(y^2)$ and $\tau(y^3)$ are arbitrary holomorphic functions of their arguments and $r_1 \neq 0$, $r_3 \neq 0$, r_2, r_4, and r_5 are arbitrary complex constants.

We also have

$$\text{(D)} \quad f(y^1) = a = \text{const} \neq 0. \tag{4.27}$$

Then we find

$$\mu = \{[3ac_1 y^2 + 3ac_2 y^3 - 2c_3 u_1 - (c_4 + c_5 + c_6)] u_1 + 2ac_7 y^2 + 2ac_8 y^3 + 4ac_3 y^2 y^3 + c_9\} y^1 + \rho(y^2, y^3, u_1)$$
$$+ [(2c_3 y^2 + \tfrac{1}{2} c_2 u_1 + c_8) u_1 + (-2ac_1 y^2 + 2c_6) y^2 + c_{10}] u_2 + [(2c_3 y^3 + \tfrac{1}{2} c_1 u_1 + c_7) u_1$$
$$+ (-2ac_2 y^3 + 2c_5) y^3 + c_{11}] u_3 + (-ac_1 y^2 - ac_2 y^3 + c_3 u_1 + c_4) u, \tag{4.28}$$

where $c_1,...,c_{11}$ are arbitrary complex constants and $\rho = \rho(y^2, y^3, u_1)$ is any solution of the following linear partial differential equation:

$$a \frac{\partial^2 \rho}{\partial u_1^2} + \frac{\partial^2 \rho}{\partial y^2 \, \partial y^3} = 0. \tag{4.29}$$

We have not been able to find explicitly the maximal group of contact transformations generated by μ of the form (4.28). However, one can find an important subgroup of this group defined by

$$c_1 = c_2 = c_3 = c_7 = c_8 = 0. \tag{4.30}$$

From (4.28) with (4.30), we get

$$\xi^1 = (c_4 + c_5 + c_6) y^1 - \frac{\partial \rho}{\partial u_1}, \quad \xi^2 = -2c_6 y^2 - c_{10},$$
$$\xi^3 = -2c_5 y^3 - c_{11}, \quad \eta = c_4 u + c_9 y_1 + \rho - u_1 \frac{\partial \rho}{\partial u_1}, \tag{4.31}$$

and, from (2.5),

$$\zeta_1 = -(c_5+c_6)u_1 + c_9, \quad \zeta_2 = (c_4+2c_6)u_2 + \frac{\partial\rho}{\partial y^2},$$
$$\zeta_3 = (c_4+2c_5)u_3 + \frac{\partial\rho}{\partial y^3}. \tag{4.32}$$

Then the contact group of transformations defined by (4.31) and (4.32) takes the form

$$y^{1'} = r_1 y^1 - \frac{\partial\varphi}{\partial u_1}, \quad y^{2'} = r_2 y^2 + r_3,$$
$$y^{3'} = r_4^2 r_1^{-2} r_2^{-1} y^3 + r_5,$$
$$u' = r_4 u + r_6\cdot\left(r_1 y^1 - \frac{\partial\varphi}{\partial u_1}\right) + r_4 r_1^{-1}\cdot\left(\varphi - u_1\frac{\partial\varphi}{\partial u_1}\right), \tag{4.33}$$
$$u'_{1'} = r_4 r_1^{-1} u_1 + r_6, \quad u'_{2'} = r_4 r_2^{-1}\cdot\left(u_2 + r_1^{-1}\frac{\partial\varphi}{\partial y^2}\right),$$
$$u'_{3'} = r_4^{-1} r_1^2 r_2\cdot\left(u_3 + r_1^{-1}\frac{\partial\varphi}{\partial y^3}\right),$$

where $r_1\neq 0$, $r_2\neq 0$, $r_4\neq 0$, r_3, r_5, and r_6 are arbitrary complex constants and $\varphi = \varphi(u_1,y^2,y^3)$ is any solution of Eq. (4.29).

Thus if $u = u(y^i)$ is any solution of Eq. (4.12) for $f(y^1)$ defined by (4.27), and

$$D_1(y^{1'}) = r_1 - u_{11}\frac{\partial^2\varphi}{\partial u_1^2}\neq 0, \tag{4.34}$$

then $u' = u'(y^{i'})$, as given by (4.33), is also a solution of this equation. Now as $\varphi = \varphi(u_1,y^2,y^3)$ is an arbitrary solution of (4.29) provided that the condition (4.34) is satisfied, one can expect that all or "almost all" solutions of Eq. (4.12) for $f(y^1)$ defined by (4.27) can be obtained from a given solution $u = u(y^i)$ with the use of the transformations (4.33).

Indeed the following theorem holds.

Theorem 4.2: Let $u = u(y^i)$ be a solution of Eq. (4.12) with (4.27) on an open neighborhood V of a point $p\in C^3$, and let

$$\left.\frac{\partial^2 u}{\partial y^{1^2}}\right|_p \neq 0.$$

Then for every point $q\in C^3$ and every local solution $u'(y^{i'})$ of (4.12) and (4.27) at q satisfying the condition

$$\left.\frac{\partial^2 u'}{\partial y^{1'^2}}\right|_q \neq 0$$

there exist an open neighborhood $V'\subset V$ of p, an open neighborhood V'' of q, and a solution $\varphi = \varphi(u_1,y^2,y^3)$ of Eq. (4.29) such that the mapping

$$\Phi:\ V'\ni(y^1,y^2,y^3)\mapsto(y^{1'},y^{2'},y^{3'})\in V'',$$
$$y^{1'} = y^1 - \left.\frac{\partial\varphi}{\partial u_1}\right|_{(y^i)},$$
$$y^{2'} = y^2 + r_3, \quad y^{3'} = y^3 + r_5, \tag{4.35}$$

is a diffeomorphism of V' onto V'', and

$$u^1(y^{i'}) = \left[u + r_6\cdot\left(y_1 - \frac{\partial\varphi}{\partial u_1}\right) + \varphi - u_1\frac{\partial\varphi}{\partial u_1}\right]_{(y^i)\in V'} \tag{4.36}$$

where r_3, r_5, and r_6 are complex constants.

Proof: First consider the case $p = q$. Let

$$b := \left.\frac{\partial u}{\partial y^1}\right|_p, \quad c := \left.\frac{\partial u'}{\partial y^{1'}}\right|_p.$$

Then the function

$$\tilde u'(y^{i'}) := u'(y^{i'}) + (b-c)y^{1'} \tag{4.37}$$

is a local solution of Eq. (4.12) for (4.27), and

$$\left.\frac{\partial\tilde u'}{\partial y^{1'}}\right|_p = \left.\frac{\partial u}{\partial y^1}\right|_p, \quad \left.\frac{\partial^2\tilde u'}{\partial y^{1'^2}}\right|_p \neq 0. \tag{4.38}$$

Then by the assumption

$$\left.\frac{\partial^2 u}{\partial y^{1^2}}\right|_p \neq 0, \left.\frac{\partial^2 u'}{\partial y^{1'^2}}\right|_p \neq 0 \Leftrightarrow \left.\frac{\partial^2\tilde u'}{\partial y^{1'^2}}\right|_p \neq 0,$$

it follows that there exist open neighborhoods $V'\subset V$ and V'' of p, and an open set $W\subset C^3$ such that the mappings

$$\Psi':\ V'\ni(y^1,y^2,y^3)\mapsto(z,y^2,y^3)\in W,$$
$$\Psi'':\ V''\ni(y^{1'},y^{2'},y^{3'})\mapsto(z',y^{2'},y^{3'})\in W,$$
$$z = \frac{\partial u(y^i)}{\partial y^1}, \quad z' = \frac{\partial\tilde u'(y^i)}{\partial y^{1'}}, \tag{4.39}$$

are diffeomorphisms of V' or V'', respectively, onto W. Define two functions on W:

$$K(z,y^2,y^3) := u(y^1(z,y^2,y^3),y^2,y^3) - zy^1(z,y^2,y^3),$$
$$K'(z',y^{2'},y^{3'}) := u'(y^{1'}(z',y^{2'},y^{3'}),y^{2'},y^{3'}) - z'y^{1'}(z',y^{2'},y^{3'}), \tag{4.40}$$

where $y^1(z,y^2,y^3)$, and $y^{1'}(z',y^{2'},y^{3'})$ are solutions of the equations

$$z = \frac{\partial u(y^i)}{\partial y^1}, \quad\text{and}\quad z' = \frac{\partial u'(y^{i'})}{\partial y^{1'}}$$

with respect to y^1 and $y^{1'}$, respectively. Then K and K' satisfy a linear partial differential equation of the type (4.29), i.e.,

$$a\frac{\partial^2 K}{\partial z^2} + \frac{\partial^2 K}{\partial y^2\,\partial y^3} = 0 \quad\text{and}\quad a\frac{\partial^2 K'}{\partial z'^2} + \frac{\partial^2 K'}{\partial y^{2'}\,\partial y^{3'}} = 0; \tag{4.41}$$

moreover,

$$\frac{\partial^2 K}{\partial z^2}\neq 0 \quad\text{and}\quad \frac{\partial^2 K'}{\partial z'^2}\neq 0 \quad\text{on } W. \tag{4.42}$$

From (4.41) one infers that there exists a function $\varphi = \varphi(z,y^2,y^3)$ on W such that

$$a\frac{\partial^2\varphi}{\partial z^2} + \frac{\partial^2\varphi}{\partial y^2\,\partial y^3} = 0 \tag{4.43}$$

and

$$K'(z,y^2,y^3) = K(z,y^2,y^3) + \varphi(z,y^2,y^3). \tag{4.44}$$

Then from (4.39), (4.40), (4.42), and (4.44) it follows that the mapping

$$\Phi = \Psi''^{-1}\circ\Psi':\ V'\ni(y^1,y^2,y^3)\mapsto(y^{1'},y^{2'},y^{3'})\in V'',$$
$$y^{1'} = y^1 - \left.\frac{\partial\varphi}{\partial u_1}\right|_{(y^i)},$$
$$y^{2'} = y^2, \quad y^{3'} = y^3, \tag{4.45}$$

is a diffeomorphism of V' onto V''.

Finally (4.37), (4.39), (4.40), (4.44), and (4.45) yield (4.36) with $r_6 = c - b$. Hence for $p = q$ our theorem holds. Assume $p \neq q$. Let $p = (y_p^1, y_p^2, y_p^3)$, $q = (y_q^1, y_q^2, y_q^3)$. Then there exist complex constants s, r_3, and r_5 such that

$$y_q^1 = y_p^1 + s, \quad y_q^2 = y_p^2 + r_3, \quad y_q^3 = y_p^3 + r_5. \tag{4.46}$$

Define an open set $V_1 \subset C^3$ $(q \in V_1)$:

$$V_1 := \{(y^1, y^2, y^3) \in C^3: \\ (y^1 - s, y^2 - r_3, y^3 - r_5) \in V\}. \tag{4.47}$$

The function $\hat{u}$: $V_1 \to C$,

$$\hat{u}: V_1 \ni (y^1, y^2, y^3) \mapsto u(y^1 - s, y^2 - r_3, y^3 - r_5), \tag{4.48}$$

is a solution of Eq. (4.12) for (4.27) on V_1, and

$$\left.\frac{\partial^2 \hat{u}}{\partial y^{1^2}}\right|_q = \left.\frac{\partial^2 u}{\partial y^{1^2}}\right|_p \neq 0.$$

Thus $\hat{u}(y^i)$ and $u'(y^{i'})$ appear to be two solutions of Eq. (4.12) for (4.27) on some neighborhoods of the point q.

One also has

$$\left.\frac{\partial^2 \hat{u}}{\partial y^{1^2}}\right|_q \neq 0, \quad \left.\frac{\partial^2 u'}{\partial y^{1'^2}}\right|_q \neq 0.$$

Therefore we can repeat our previous considerations and we get (4.35) and (4.36). The proof is completed. ■

Concluding, one can say that any solution $u'(y')$ of Eq. (4.12) with (4.27) on a sufficiently small open set of C^3, provided that $\partial^2 u'/\partial y^{1^2} \neq 0$ is generated by a given local solution $u = u(y^i)$, $\partial^2 u^i/\partial y^{1^2} \neq 0$, with the use of a suitable transformation (4.33) for $r_1 = r_2 = r_4 = 1$. This is an obvious consequence of the fact that Eq. (4.12) for $f(y^1) = a = \text{const} \neq 0$ and $u_{11} \neq 0$ can be linearized by the contact transformation[18–20] according to (4.39)–(4.41).

[*Remark:* Analogous conclusions can be obtained for the case $u_{11} = 0$. However, in this case,

$$u(y^1, y^2, y^3) = v(y^2, y^3) y^1 + w(y^2, y^3),$$

where $v(y^2, y^3)$ satisfies the following equation:

$$v_2 v_3 - a = 0,$$

and $w(y^2, y^3)$ is an arbitrary holomorphic function of y^2 and y^3. Thus the problem appears to be two dimensional and it is not interesting for our purposes.]

Finally we have

$$\text{(E)}\ f(y^1) = 0. \tag{4.49}$$

Then

$$\mu = \sigma(u_1) y^1 + \omega(y^2, u_1) + \tau(y^3, u_1) + \alpha(y^2, u_1, u_2) \\ + \beta(y^3, u_1, u_3) + \gamma(u_1) u, \tag{4.50}$$

where $\sigma(u_1)$, $\omega(y^2, u_1)$, $\tau(y^3, u_1)$, $\alpha(y^2, u_1, u_2)$, $\beta(y^3, u_1, u_3)$, and $\gamma(u_1)$ are arbitrary holomorphic functions of their arguments.

We have not succeeded in finding the general group of contact transformations defined by μ of the form (4.50) but we can find an important subgroup of this group. It is exactly of the form (4.33) with $\varphi = \varphi(u_1, y_2, y_3)$ now an arbitrary solution of the following linear partial differential equation:

$$\frac{\partial^2 \varphi}{\partial y^2 \partial y^3} = 0, \tag{4.51}$$

i.e., Eq. (4.29) for $a = 0$. Therefore

$$\varphi(u_1, y^2, y^3) = \delta(u_1, y^2) + \varepsilon(u_1, y^3), \tag{4.52}$$

where $\delta(u_1 y^2)$ and $\varepsilon(u_1, y^3)$ are holomorphic functions of their arguments. Then Theorem 4.2 holds true for the present case provided $a = 0$. The analysis of Eq. (4.12) is closed.

Up to now, to the best of our knowledge, only the cases (B) and (D) are employed in complex relativity and the gravitational instanton theory. The specialization of the results obtained for these cases to the gravitational instanton theory can be achieved by imposing the following restrictions:

(B) a, b, c_1, c_2, r_1, r_2 real constants,

$$y^1 \text{ real}, \quad y^2, y^3 \text{ complex}, \quad \overline{y^2} = y^3, \quad u \text{ real},$$

$$\overline{\rho(y^2)} = \omega(y^3), \quad \overline{\gamma(y^2)} = \delta(y^3),$$

$$\overline{\varphi(y^2)} = \psi(y^3), \quad \overline{\sigma(y^2)} = \tau(y^3);$$

(D) $a, c_3, c_4, c_9, r_1, r_4, r_6$ real constants,

$$\overline{c_1} = c_2, \quad \overline{c_5} = c_6, \quad \overline{c_7} = c_8, \quad \overline{c_{10}} = c_{11},$$

$$\overline{r_3} = r_5, \quad r_2 \overline{r_2} = r_4^2 r_1^{-2},$$

$$y^1 \text{ real}, \quad y^2, y^3 \text{ complex}, \quad \overline{y^2} = y^3, \quad u \text{ real},$$

$$\rho(y^2, y^3, u^1),\ \varphi(u_1, y^2, y^3) \text{ real}.$$

Then Theorem 4.2 remains (*mutatis mutandis*) valid. Consequently all Gibbons–Hawking metrics can be generated from any given Gibbons–Hawking metric by means of transformations (4.33) for $r_1 = r_2 = r_4 = 1$. An important conclusion concerning case (B) can be readily derived from (4.23). As we have pointed out, case (B) for $a = b = 1$ corresponds to the self-dual, Ricci-flat gravitational instanton admitting the "rotational" Killing vector field or to the heaven of "case III." The transformation of the corresponding metric caused by (4.23) appears to be a composition of some coordinate transformation and the conformal transformation of the constant conformal factor $e^{(1/2)r_1}$. Thus statements (i) and (ii) of Sec. II hold (*mutatis mutandis*) true in the present case.

[1]S. Lie, *Differentialgleichungen* (Chelsea, New York, 1967).
[2]G. Birkhoff, *Hydrodynamics* (Princeton U. P., Princeton, NJ, 1950).
[3]L. V. Ovsiannikov, *Group Analysis of Differential Equations* (Academic, New York, 1982).
[4]G. W. Bluman and J. D. Cole, *Similarity Methods for Differential Equations* (Springer, New York, 1974).
[5]R. L. Anderson and N. H. Ibragimov, *Lie–Bäcklund Transformations in Applications* (SIAM, Philadelphia, 1979).
[6]N. H. Ibragimov, *Gruppy Preobrazovanii v Matematicheskoi Fizike* (Nauka, Moscow, 1983) (in Russian).
[7]J. F. Plebański, J. Math. Phys. **16**, 2395 (1975).
[8]C. P. Boyer, I. D. Finley, III, and J. F. Plebański, "Complex General Relativity $\mathcal{H}$ and $\mathcal{HH}$ Spaces—A Survey," Com. Tec. Inst. Inv. Mat. Apl. Sist. UNAM No. 174, 1978.
[9]E. J. Flaherty, *Hermitian and Kählerian Geometry in Relativity* (Springer, Berlin, 1976).
[10]M. F. Atiyah, N. J. Hitchin, and I. M. Singer, Proc. R. Soc. London, Ser. A **362**, 425 (1978).
[11]S. Hacyan, Phys. Lett. A **75**, 23 (1979).
[12]G. W. Gibbons and C. N. Pope, Commun. Math. Phys. **66**, 267 (1979).

[13]M. Przanowski and B. Broda, Acta Phys. Pol. B **14**, 637 (1983).
[14]M. Przanowski, Lett. Acta Phys. Pol. B **14**, 625 (1983).
[15]M. Przanowski and B. Baka, Gen. Relativ. Gravit. **16**, 797 (1984).
[16]S. W. Hawking, Phys. Lett. A **60**, 81 (1977).
[17]G. W. Gibbons and S. W. Hawking, Commun. Math. Phys. **66**, 291 (1979).
[18]J. D. Finley, III and J. F. Plebański, J. Math. Phys. **20**, 1938 (1979).
[19]C. P. Boyer and J. D. Finley, III, J. Math. Phys. **23**, 1126 (1982).
[20]M. Przanowski and S. Białecki, Acta Phys. Pol. B **18**, 879 (1987).
[21]N. H. Ibragimov, *Gruppovye svoistva nekotoryh differencialnyh uravnenii* (Nauka, Novosibirsk, 1967).
[22]Pham Mau Quan, C. R. Acad. Sci. Paris Ser. A **285**, 1081 (1977).
[23]J. F. Pommaret, *Systems of Partial Differential Equations and Lie Pseudogroups* (Gordon and Breach, New York, 1978).
[24]J. D. Gegenberg and A. Das, Gen. Relativ. Gravit. **16**, 817 (1984).

Form perturbations of the Laplacian on $L^2(\mathbb{R})$ by a class of measures

A. Tip
FOM-Institute for Atomic and Molecular Physics, Kruislaan 407, 1098 SJ Amsterdam, The Netherlands

(Received 30 September 1987; accepted for publication 23 August 1989)

A class of form perturbations of the Laplacian $-\partial_x^2$ on $L^2(\mathbb{R}, dx)$, defined in terms of regular, in general, complex, Borel measures, satisfying a certain growth condition, is studied. The domain of the associated closed operator H is characterized and the special case of measures with finite total measure is investigated. A unicity theorem for the eigenfunctions of H is derived.

I. INTRODUCTION

One-dimensional systems play an important role in illustrating concepts from both quantum mechanics and operator theory. Thus in almost any textbook on quantum mechanics the reader is confronted with the one-dimensional quantum mechanical motion of a particle generated by a Hamiltonian of the type

$$H = -\partial_x^2 + V(x), \tag{1.1}$$

acting in $\mathscr{H} = L^2(\mathbb{R}, dx)$, where the potential $V(x)$ is a square well, a step function, or a Dirac δ function. On the other hand, such systems are a convenient means to study operator theory as witnessed by the monograph by Schechter,[1] which is completely devoted to the former. Since H has to generate a unitary time evolution it must possess a self-adjoint extension. Since the Laplacian $-\partial_x^2$ is essentially self-adjoint on the Schwartz space $\mathscr{S} \subset \mathscr{H}$ with closure $p^2 = H_0 = (-\partial_x^2)^c$ (domain $\mathscr{D}_0$), it follows by a well known theorem[2,3] that for symmetric potentials $V(x)$ with domain containing $\mathscr{D}_0$ and which are $(p^2, 1-\varepsilon)$-bounded, $H_0 + V$ is self-adjoint with domain $\mathscr{D}_0$. It is also known that δ functions and general $V(x) \in L^1(\mathbb{R}, dx)$ are not in this class. They can, however, be handled by quadratic form techniques[2,3] (for a different approach, see Ref. 4). Such techniques have been considered by a number of authors, especially in the three-dimensional case.[5,6] Quite general results have been obtained in the N-dimensional case by Herbst and Sloan,[7] who consider perturbations of H_0 (which can be more general than considered here) by the sum of a non-negative potential and a small form perturbation.

Observing that both δ and L_1 potentials define a measure $d\mu(x) = V(x)dx$, it makes sense to investigate the situation where the Laplacian is perturbed by measures, thus unifying the various cases mentioned above. Thus we consider its form perturbations by a class $\mathfrak{M}$ of regular, in general, complex, Borel measures μ, characterized by a growth condition stating that, on the average, $|\mu|([a,b])$ grows at most linearly with the length $l = b - a$ of the interval $[a,b]$. This includes cases such as infinite or semi-infinite Kronig–Penney lattices. In Sec. II we give a general expression for $[z-H]^{-1}$, $z \in \rho(H)$ (the resolvent set of H), and give a characterization of the functions in the domain $\mathscr{D}(H)$ of H. Decomposing $\mu \in \mathfrak{M}$ into its atomic, singularly continuous, and absolutely continuous parts, $\mu = \mu^{at} + \mu^{sc} + \mu^{ac}$, we have the expected situation of boundary conditions in the points of $C(\mu^{at})$ [$C(\mu)$ is the set on which μ is concentrated], whereas μ^{ac} leads to a multiplicative perturbation $V(x)$, $d\mu^{ac}(x) = V(x)dx$, $V(x) \in L^1_{loc}(\mathbb{R}, dx)$. Singularly continuous measures, however, make their presence felt in a more subtle way. (Here and in the following, phrases such as absolutely continuous, singularly continuous, and almost everywhere are with respect to Lebesgue measure.) In Sec. III measures μ with finite total measure, $\|\mu\| = |\mu|(\mathbb{R}) < \infty$, are considered. Now $[z\text{-}H]^{-1} - [z\text{-}H_0]^{-1}$, $z \in \rho(H) \cap \rho(H_0)$, is a trace class operator, leading to a scattering situation possessing asymptotic completeness for self-adjoint H (real μ). In Sec. IV we derive a unicity theorem for the eigenfunctions of H. One of its consequences is that for μ real, concentrated on a finite of semi-finite interval, H has no non-negative eigenvalues. A similar case, related to a quantum mechanical tunneling situation, is briefly mentioned in the discussion section. The results concerning operator and form techniques, used in the sequel, can be found in Refs. 2 and 3. For the various measure-theoretic notions, Ref. 8 is a convenient reference.

II. MEASURES AS FORM PERTURBATIONS

We start by introducing some notation. As mentioned earlier, we are considering form perturbations of $H_0 = p^2$ (domain $\mathscr{D}_0$), the closure of $-\partial_x^2$ on $\mathscr{S}(\mathbb{R}) \subset \mathscr{H} = L^2(\mathbb{R}, dx)$ [inner product (f,g), norm $\|f\|$]. The closed quadratic form associated with H_0 is

$$\Phi_0(f,g) = (H_0^{1/2}f, H_0^{1/2}g) = (pf, pg),$$
$$f,g \in \mathscr{D}(H_0^{1/2}) = \mathscr{D}(p) = \mathscr{D},$$

where p is the momentum operator [the closure of $-i\partial_x$ on $\mathscr{S}(\mathbb{R})$]. We recall that with the resolvents of p and p^2 are associated the integral kernels [$\theta(x)$ is the unit step function; $\theta(x) = 1$, $x \geqslant 0$, zero, otherwise]

$$\langle x|[z-p]^{-1}|y\rangle = \begin{cases} -i\theta(x-y)\exp[iz(x-y)], & \operatorname{Im} z > 0, \\ i\theta(y-x)\exp[i\sqrt{z}(x-y)], & \operatorname{Im} z < 0, \end{cases} \tag{2.1}$$

$$\langle x|[z-p^2]^{-1}|y\rangle = [2i\sqrt{z}]^{-1}\exp[i\sqrt{z}|x-y|],$$
$$z \in \mathbb{C} \setminus [0,\infty).$$

Noting that any $f \in \mathscr{D}(p)$, its Fourier transform being an L^1 function, has a continuous version vanishing in infinity, denoted by $\langle x|f\rangle$ (and its complex conjugate by $\langle f|x\rangle$), we

 0022-2488/90/020308-08$03.00

see that $\langle x|f\rangle \in C_0 = C_0(\mathbb{R})$, the algebra of continuous functions, vanishing in infinity. Now, C_0 equipped with the sup-norm is a Banach algebra and its dual, C_0^*, is the space of regular Borel measures μ with finite total measure, the norm being $\|\mu\| = |\mu|(\mathbb{R})$ (for these facts, see Ref. 8). Thus

$$\Phi_1(f,g) = \int d\mu(x)\langle g|x\rangle\langle x|f\rangle, \quad f,g \in \mathscr{D}, \quad \mu \in C_0^*,$$

is well defined (integrals, unless specified differently, are over $\mathbb{R}$) so that

$$\Phi(f,g) = \Phi_0(f,g) + \Phi_1(f,g) \tag{2.2}$$

is also properly defined and a simple estimate shows that Φ_1 is Φ_0-bounded with zero relative bound so that Φ defines an m-sectorial operator H with $\mathscr{D}(H) \subset \mathscr{D}$ and $(Hf,g) = \Phi(f,g), f \in \mathscr{D}(H)$. In particular, H is closed and, in case μ is real, self-adjoint and bounded from below.[2] Replacing μ by Lebesgue measure in (2.2) we see that Φ_1 perturbs H_0 through the addition of a constant. This suggests the extension of our class of measures by incorporating measures that, on the average, grow as Lebesgue measure.

Definition 2.1: Let μ be a regular Borel measure. Then μ is said to belong to the class $\mathfrak{M}$ if there exist positive constants κ and l such that $|\mu|([a\text{-}b]) \leqslant \kappa(b,a)$ for every finite interval with length $b\text{-}a \geqslant l$. A scaling argument shows that we can take $l = 1$ without loss of generality as we shall do in the sequel.

Remark: The measure μ can be complex-valued with infinite total measure, which deviates from the standard definition. Note, however, that μ, restricted to a finite interval, has finite total measure. In the sequel $d\mu$ is always weighed in such a way that the weighed measure has finite total measure.

Proposition 2.2: Let $\mu \in \mathfrak{M}$. Then (2.2) defines a closed m-sectorial operator H with domain $\mathscr{D}(H) \subset \mathscr{D}$. For real μ, H is self-adjoint and bounded from below.

Proof: We show that Φ_1 is Φ_0-bounded with zero relative bound. Let $f \in \mathscr{D}(p)$ so that $f = [ia\text{-}p]^{-1}\varphi$, $a > 0$, $\varphi \in \mathscr{H}$. Proceeding formally, using (2.1), we have

$$\begin{aligned}\Phi_1(f,f) &= \int d\mu(y)|\langle y|f\rangle|^2 \\ &= \int d\mu(y)\int dx\, dx'\, \theta(y-x)\theta(y-x') \\ &\quad \times \exp[-a(2y-x-x')]\varphi(x)\bar{\varphi}(x') \\ &= \int dx M(x,a)h(x),\end{aligned}$$

where

$$\begin{aligned}h(x) &= \int_0^\infty dx' \exp[-ax'] \\ &\quad \times \{\varphi(x)\bar{\varphi}(x-x') + \varphi(x-x')\bar{\varphi}(x)\}, \\ M(x,a) &= \int d\mu(y)\theta(y-x)\exp[-2a(y-x)].\end{aligned}$$

Now $h \in L^1(\mathbb{R}, dx)$ with $\|h\|_1 < 2a^{-1}\|\varphi\|^2$ and, writing

$$[x,\infty) = \bigcup_{k=0}^{\infty}[x+k, x+k+1],$$

we have

$$|M(x,a)| < \kappa[1 - \exp[-2a]]^{-1},$$

so that our formal manipulations are justified and

$$\begin{aligned}|\Phi_1(f,f)| &< 2\kappa a^{-1}[1-\exp[-2a]]^{-1}\|\varphi\|^2 \\ &= 2\kappa a^{-1}[1-\exp[-2a]]^{-1}\{\Phi_0(f,f) \\ &\quad + a^2\|f\|^2\}.\end{aligned}$$

□

Remark: We note that $\mathfrak{M}$ contains lattices of δ potentials (Kronig–Penney lattices) but also atomic measures where $C(\mu^{\text{at}})$ clusters locally, provided the strength of the δ potentials decays sufficiently fast.

Proposition 2.3: Let H be as in Proposition 2.2. For $\operatorname{Im}\sqrt{z} > 2\kappa$, we have $z \in \rho(H)$ and

$$\begin{aligned}&[z-H]^{-1} \\ &\quad = [\sqrt{z}+p]^{-1}[1+K(z)]^{-1}[\sqrt{z}-p]^{-1} = R(z),\end{aligned} \tag{2.3}$$

where the bounded operator $K(z)$, defined by the integral kernel

$$\begin{aligned}\langle x|K(z)|x'\rangle &= \int d\mu(y)\exp[i\sqrt{z}(x+x'-2y)] \\ &\quad \times \theta(x-y)\theta(x'-y),\end{aligned} \tag{2.4}$$

is analytic in $z \in \mathbb{C}\setminus[0,\infty)$ and obeys

$$\|K(z)\| < 2\kappa(\operatorname{Im}\sqrt{z})^{-1}[1-\exp[-2\operatorname{Im}\sqrt{z}]^{-1}]. \tag{2.5}$$

Proof: (a) Let $\sqrt{z} = u + iv$. For $x > x'$, we have

$$\begin{aligned}&|\langle x|K(z)|x'\rangle| \\ &\quad < \int d|\mu|(y)\exp[-v(x+x'-2y)]\theta(x'-y) \\ &\quad = \int d|\mu|(y)\chi_{(-\infty,x']}(y) \\ &\qquad \times \exp[-2v(x'-y)]\exp[-v(x-x')] \\ &\quad < \kappa[1-\exp[-2v]]^{-1}\exp[-v(x-x')],\end{aligned}$$

and similarly for $x < x'$ [$\chi_A(y)$ is the characteristic function of the set A] with the result

$$\begin{aligned}&|\langle x|K(z)|x'\rangle| \\ &\quad \leqslant \kappa[1-\exp[-2v]]^{-1}\exp[-v|x-x'|] \\ &\quad = 2\kappa v[1-\exp[-2v]]^{-1}\langle x|[p^2+v^2]^{-1}|x'\rangle,\end{aligned}$$

from which (2.5) follows. The analyticity statement follows by applying Morera's and Fubini's theorems to $(K(z)f,g)$, $f,g \in \mathscr{H}$.

(b) The existence of $K(z)$ now being established, we note that a suggestive way of writing $K(z)$ is

$$K(z) = -\int d\mu(y)[\sqrt{z}-p]^{-1}|y\rangle\langle y|[\sqrt{z}+p]^{-1}. \tag{2.6}$$

This expression suggests the relation

$$\begin{aligned}K(z_1) &= [\sqrt{z_1}-p]^{-1}[\sqrt{z_2}-p]K(z_2)[\sqrt{z_2}+p] \\ &\quad \times [\sqrt{z_1}+p]^{-1},\end{aligned}$$

which is readily vertified by actual computation. Using this result we find that $R(z)$ obeys the resolvent equation $R(z_1)-R(z_2)=(z_2-z_1)R(z_1)R(z_2)$, which implies that the range of $R(z)$ is z-independent. In addition, (2.3) shows that $R(z)$ and $R^*(z)$ have empty null spaces, so that $R(z)^{-1}=z-H$ exists with H closed, densely defined. For $f,g\in\mathscr{D}$, we have

$$(R(z)^{-1}f,g)=(\{1+K(z)\}(\sqrt{z}+p)f,(\overline{\sqrt{z}}-p)g)$$
$$=z(f,g)-\Phi(f,g),$$

so that H coincides with our earlier definition. □

Remark: We shall say that a function $f(x)$ is contained in BV_{loc} if its restriction to a finite interval is of bounded variation.

Theorem 2.4: Let H be as in Proposition 2.2. Then we have the following:

(a) In case $\mathscr{D}(H)=\mathscr{D}(H_0)=\mathscr{D}_0$, μ is absolutely continuous: $d\mu(x)=V(x)dx$, $V(x)\in L^2_{\text{loc}}(\mathbb{R},dx)$.

(b) Every $f\in\mathscr{D}(H)$ has a continuous version $\langle x|f\rangle$, continuously differentiable in $\mathbb{R}\backslash C(\mu^{\text{at}})$. In each $x\in C(\mu^{\text{at}})$, $\langle x|f\rangle$ possesses left and right derivatives, their difference being equal to $\mu(\{x\})\langle x|f\rangle$ and both being contained in BV_{loc}.

(c) Let D_x be the Lebesgue derivative and let f be as in (b). Then for almost every x,

$$(Hf)(x)=\{-D_x\,\partial_x+V(x)\}\langle x|f\rangle$$
$$=\{-D_x^2+V(x)\}\langle x|f\rangle, \tag{2.7}$$

where $V(x)$ is defined through $d\mu^{\text{ac}}(x)=V(x)dx$.

(d) Let f be as in (b) and let $(\partial_x\langle x|f\rangle)^s$ be the singular part of $\partial_x\langle x|f\rangle$ [the latter is not defined in $x\in C(\mu^{\text{at}})$ but we can take either the left or right derivative of $\langle x|f\rangle$; its singular part is defined in terms of the associated singular measure]. Then, for each $h(x)\in\mathscr{S}$ ($\mu^s=\mu^{\text{sc}}+\mu^{\text{at}}$),

$$\int dx(\partial_x\langle x|f\rangle)^s\,\partial_x\,\overline{h(x)}=\int d\mu^s(x)\langle x|f\rangle\,\overline{h(x)}. \tag{2.8}$$

We start with a lemma.

Lemma 2.5: Let $\nu\in\mathfrak{M}$, $\gamma=\text{Re}\,\zeta>0$, and $f\in C_0(\mathbb{R})$. Then

$$g(x)=\int d\nu(y)\exp[i\zeta|x-y|]f(y)$$

exists for each $x\in\mathbb{R}$, is bounded in absolute value by an x-independent constant, and tends to zero as $|x|\to\infty$.

Proof: For each $\varepsilon>0$, $\exists\, y_0>1$ such that $|f(y)|<\varepsilon$, for $|y|>y_0$. Thus

$$|g(x)|\leqslant\varepsilon\int d|\nu|(y)\chi_{(y_0,\infty)}(|y|)\exp[-\gamma|x-y|]$$
$$+\|f\|_\infty\int d|\nu|(y)\chi_{(0,y_0)}(|y|)\exp[-\gamma|x-y|]$$
$$\leqslant\varepsilon\int d|\nu|(y)\chi_{(y_0,\infty)}(|y|)\exp[-\gamma|x-y|]$$
$$+2\kappa\|f\|_\infty\max_{|y|\in[0,y_0]}\exp[-\gamma|x-y|].$$

The second term has the needed behavior and it remains to consider the first. For $|x|\leqslant y_0$, the first term is properly behaved and it remains to consider its properties for $|x|>y_0$. Thus let $x>y_0$ (the case $x<-y_0$ goes similarly). Now, using $\nu\in\mathfrak{M}$,

$$\int d|\nu|(y)\chi_{(-\infty,-y_0)}(y)\exp[-\gamma|x-y|]$$
$$=\int d|\nu|(y)\chi_{(-\infty,-y_0)}(y)\exp[-\gamma y]\exp[-\gamma x]$$
$$\leqslant k\exp[-\gamma x],$$

where $k>0$ is a constant. Second (n_1 is the first integer such that $y_0+n_1\geqslant x$ and $k'>0$ a constant),

$$\int d|\nu|(y)\chi_{(y_0,\infty)}(y)\exp[-\gamma|x-y|]$$
$$=\int d|\nu|(y)\chi_{(y_0,x}(y)\exp[\gamma y]\exp[-\gamma x]$$
$$+\int d|\nu|(y)\chi_{(x,\infty)}(y)\exp[-\gamma(x-y)]$$
$$\leqslant\kappa\sum_{n=0}^{n_1}\exp[\gamma(y_0+n)]\exp[-\gamma x]$$
$$+\kappa\sum_{n=0}^{\infty}\exp[-\gamma(x+n)]\exp[\gamma x\leqslant k'.]$$

Thus we have assembled all ingredients needed for the validity of the lemma. □

Proof of Theorem 2.4: (a) Let $\mathscr{D}(H)=\mathscr{D}(H_0)=\mathscr{D}_0$ and let $f\in\mathscr{D}_0$ be such that $f(y)=1$ on the finite interval $[a,b]$, $b>a$. We denote

$$V(y)=(Hf)(y)-(H_0f)(y),\ y\in[a,b].$$

Then for $g\in\mathscr{D}$ with support in (a,b),

$$\int_a^b dy\,V(y)\langle y|g\rangle=\int d\mu(y)\langle y|g\rangle$$

and by a limiting argument

$$\int_a^x dy\,V(y)=\mu((a,x)),\quad x\in(a,b).$$

Since $V\in L^2_{\text{loc}}$, $V\in L^1_{\text{loc}}$ and $d\mu(y)=V(y)dy$.

(b) Since $\mathscr{D}(H)$ is contained in $\mathscr{D}(p)$, every $f\in\mathscr{D}(H)$ has a continuous version $\langle x|f\rangle$. Writing $f=[z-H]^{-1}g$, $z\in\rho(H)$, $g\in\mathscr{H}$, and noting that

$$[z-H]^{-1}=[z-H_0]^{-1}-[\sqrt{z}+p]^{-1}K(z)$$
$$\times[1+K(z)]^{-1}[\sqrt{z}-p]^{-1}$$
$$=[z-H_0]^{-1}-[\sqrt{z}+p]^{-1}K(z)$$
$$\times[\sqrt{z}+p][z-H]^{-1}, \tag{2.9}$$

we have

$$\langle x|f\rangle=\langle x|[z-H_0]^{-1}g\rangle$$
$$-\langle x|[\sqrt{z}+p]^{-1}K(z)[\sqrt{z}+p]f\rangle$$
$$=\langle x|f_1\rangle-\langle x|f_2\rangle, \tag{2.10}$$

where $\langle x|f_1\rangle\in AC^2(\mathbb{R})$ [since $f_1\in\mathscr{D}(H_0)$] and

$$\langle x|f_2\rangle = [2i\sqrt{z}]^{1}\int d\mu(y)\exp[i\sqrt{z}|x-y|]\langle y|f\rangle. \tag{2.11}$$

Its atomic contribution is

$$\langle x|f_2\rangle^{\rm at} = [2i\sqrt{z}]^{-1}\sum_{x_j\in C(\mu^{\rm at})}\mu(\{x_j\}) \times\exp[i\sqrt{z}|x-x_j|]\langle x_j|f\rangle. \tag{2.12}$$

For $x \notin C(\mu^{\rm at})$, the kth derivative of the jth term in (2.10) equals

$\mu(\{x_j\})[i\sqrt{z}\,{\rm sgn}(x-x_j)]^k\cdot\exp[i\sqrt{z}|x-x_j|]\langle x_j|f\rangle.$

Applying Lemma 2.5 with $\nu=\mu^{\rm at}$, we find that (2.10) and the corresponding series for the derivatives converge absolutely, are bounded by an x-independent constant, and vanish as $|x|\to\infty$. Thus $\langle x|f_2\rangle^{\rm at}$ is arbitrarily often differentiable in $\mathbb{R}\setminus C(\mu^{\rm at})$ and possesses a left and right derivative in each $x_j\in C(\mu^{\rm at})$, their difference being equal to $\mu^{\rm at}(\{x_j\})\langle x_j|f\rangle$. In particular, $\partial_x^2\langle x|f_2\rangle^{\rm at} = -zf_2^{\rm at}(x)$, for $x\in\mathbb{R}\setminus C(\mu^{\rm at})$, and it follows that $\partial_x\langle x|f_2\rangle^{\rm at}$ can be extended to a function contained in $BV_{\rm loc}$ (the left and right derivatives of $\langle x|f_2\rangle^{\rm at}$ are such extensions).

Next we consider the continuous contribution

$$\langle x|f_2\rangle^c = [2i\sqrt{z}^{-1}\{\exp[i\sqrt{z}x]\nu_1((-\infty,x)) + \exp[-i\sqrt{z}x]\nu_2([x,\infty))\},$$

where

$$\nu_1((-\infty,x)) = \int d\mu^c(y)\chi_{(-\infty,x)}(y)\exp[-i\sqrt{z}y]\langle y|f\rangle,$$

$$\nu_2([x,\infty)) = \int d\mu^c(y)\chi_{[x,\infty)}(y)\exp[i\sqrt{z}y]\langle y|f\rangle.$$

It is easily verified that $\langle x|f_2\rangle^c$ is differentiable, its (continuous) derivative being given by

$$\partial_x\langle x|f_2\rangle^c = \tfrac{1}{2}\{\exp[i\sqrt{z}x]\nu_1((-\infty,x)) - \exp[-i\sqrt{z}x]\nu_2([x,\infty))\}$$
$$= \tfrac{1}{2}\int d\mu^c(y){\rm sgn}(x-y) \times\exp[i\sqrt{z}|x-y|]\langle y|f\rangle,$$

which vanishes as $|x|\to\infty$, according to Lemma 2.5.

The same result holds for $\langle x|f_1\rangle$ [μ^c is replaced by Lebesgue measure and $\langle y|f\rangle$ by $g(y)$].

We note in passing that, for $x\in\mathbb{R}\setminus C(\mu^{\rm at})$,

$$\partial_x\langle x|f_2\rangle = \tfrac{1}{2}\int dy\,{\rm sgn}(x-y)\exp[i\sqrt{z}|x-y|]g(y) - \tfrac{1}{2}\int d\mu(y){\rm sgn}(x-y)\exp[i\sqrt{z}|x-y|]\langle y|f\rangle. \tag{2.13}$$

(c) Since ν_1 and ν_2 are Borel measures, their derivatives with respect to Lebesgue measure exist and equal $\pm\exp[\mp i\sqrt{z}x]\langle x|f\rangle$ for almost every x so that

$$D_x(\partial_x\langle x|f_2\rangle^c) = -z\langle x|f_2\rangle^c + V(x)\langle x|f\rangle,$$

$V(x)$ being defined by $d\mu^{\rm ac}(x) = V(x)dx$. Similarly

$$D_x(\partial_x\langle x|f_1\rangle) = -z\langle x|f_1\rangle + g(x),$$

for almost every x. Combining results $\{z+D_x\partial_x - V(x)\}\langle x|f\rangle = g(x)$ and, since $g=(z-H)f$,

$$(Hf)(x) = \{-D_x\partial_x + V(x)\}\langle x|f\rangle = \{-D_x^2 + V(x)\}\langle x|f\rangle, \tag{2.14}$$

for almost every x.

(d) Since $\partial_x\langle x|f\rangle$ has an extension in $BV_{\rm loc}$ we can split it into its singular and absolutely continuous parts and write, assuming $\partial_x\langle x|f\rangle$ to exist in x_0, for a.e. x,

$$\partial_x\langle x|f\rangle = (\partial_x\langle x|f\rangle)^s + (\partial_x\langle x|f\rangle)^{\rm ac}_{x=x_0} + \int_{x_0}^x dy\,D_y\,\partial_y\langle y|f\rangle.$$

Then, for arbitrary $h\in\mathscr{S}$,

$$\Phi(f,h) = -\int dx(\partial_x\langle x|f\rangle)^s\partial_x\overline{h(x)} - \int dx\{D_x\,\partial_x\langle x|f\rangle\}\overline{h(x)} = (Hf,h).$$

Hence, using (2.14),

$$\int dx(\partial_x\langle x|f\rangle)^s\partial_x\overline{h(x)} = \int d\mu^s(x)\langle x|f|\rangle\overline{h(x)},$$

which is (2.8). □

Theorem 2.6: Let $f(x)$ be an absolutely continuous square integrable function with square integrable derivative $\partial_x f(x) = D_x f(x)$, the latter being the restriction to its domain of existence of a function $\varphi\in BV_{\rm loc}$. In addition, let its singular part $(\partial_x f)^s(x)$ satisfy

$$((\partial_x f)^s,\partial_x h) = \int d\mu^s(x)f(x)\overline{h(x)},\quad\forall h\in\mathscr{S}. \tag{2.15}$$

Then $f\in\mathscr{D}(H)$.

Proof: Since, according to our assumptions,

$$\Phi(f,h) = (\{-D_x^2 f + V(x)\}f,h),\quad\forall h\in\mathscr{S},$$

and since $\mathscr{S}$ is a core for Φ, $f\in\mathscr{D}(H)$. □

Remarks: Part (a) of Theorem 2.4 shows that in case the singular part of μ is nonzero the domains of H and H_0 do not coincide. Part (b) expresses the well known fact that δ potentials are equivalent to boundary conditions [the jump condition on $\partial_x\langle x|f\rangle$ in the points of $C(\mu^{\rm at})$]. We also note that in case $\mu^{\rm at}$ vanishes, $\partial_x\langle x|f\rangle$ is continuous but is not absolutely continuous for nonzero $\mu^{\rm sc}$. In fact, we see that singular continuous measures do not show up as a boundary value perturbation of $-\partial_x^2$ nor as a multiplicative operator; they constitute a separate class of local perturbations of the Laplacian.

It is clear from Theorem 2.4 that the conditions of Theorem 2.6 are both necessary and sufficient for f to be contained in $\mathscr{D}(H)$. Note that (2.15) implies that the only discontinuities in φ are in the points x_j of $C(\mu^{\rm at})$, where φ jumps with an amount $\mu^{\rm at}(x_j)\langle x_j|f\rangle$, but that it only makes a more implicit statement about the singular continuous part of $\partial_x f$ (in terms of its weak derivative).

III. THE CASE OF FINITE TOTAL MEASURE

Throughout this section we assume that $\|\mu\| < \infty$. Then (2.4) defines a Hilbert–Schmidt operator with Schmidt norm $\|K(z)\|_2 \leqslant \|\mu\|/(2\,\mathrm{Im}\sqrt{z})$. In fact, $K(z)$ is even trace class, since, writing $\sqrt{z} = u + iv$, $u \in \mathbb{R}$ $v > 0$,

$$K(z) = -\exp[iux]\int d\nu(y)\,[p - iv]^{-1}|y\rangle \times \langle y|\,[p + iv]^{-1}\exp[iux]$$

$$= -\exp[iux]\left\{\sum_{j=1}^{4}(i)^j\int d\nu_j(y)\,[p-iv]^{-1}|y\rangle \times \langle y|\,[p+iv]^{-1}\right\}\exp[iux], \tag{3.1}$$

where

$$d\nu(y) = \exp[-2iuy]d\mu(y) = \sum_{j=1}^{4}(i)^j d\nu_j(y),$$

with ν_j non-negative, $\|\nu_1\| + \|\nu_3\| \leqslant \|\mu\|$, $\|\nu_2\| + \|\nu_4\| \leqslant \|\mu\|$. Now each

$$\int d\nu_j(y)\,[p - iv]^{-1}|y\rangle\langle y|\,[p+iv]^{-1}$$

defines a non-negative trace class operator with trace norm $\leqslant \|\nu_j\|/(2v)$. Thus we have the following proposition.

Proposition 3.1: Let $\|\mu\| < \infty$. Then $K(z)$, $z \in \mathbb{C}\setminus[0,\infty)$, is a trace class operator with trace norm obeying $\|K(z)\|_1 \leqslant \|\mu\|/(\mathrm{Im}\sqrt{z})^{-1}$ [$\leqslant \|\mu\|/(2\,\mathrm{Im}\sqrt{z})$ in case μ is real] and operator norm obeying $\|K(z)\| \leqslant \|\mu\|/(2\,\mathrm{Im}\sqrt{z})$ (since the Schmidt norm majorizes the operator norm). It now follows form (2.7) that $[z - H]^{-1} - [z - H_0]^{-1}$, $z \in \rho(H) \cap \rho(H_0)$, is trace class and we have by some well known results (see Ref. 2, Chaps. IV and X, and Ref. 9, Theorem VI-14) the following corollary.

Corollary 3.2: The essential spectra of H and H_0 coincide, $\sigma_{\mathrm{ess}}(H) = \sigma_{ess}(H_0) = [0,\infty)$; $[z - H]^{-1}$ is analytic outside $\mathbb{C}\setminus[0,\infty)$, except possibly for a set $\mathscr{C}$ of poles of finite multiplicity that can only accumulate in points of $[0,\infty)$. In case μ is real, so that H is self-adjoint, the absolutely continuous spectra of H and H_0 coincide, $\sigma_{\mathrm{ac}}(H) = \sigma_{\mathrm{ac}}(H_0) = [0,\infty)$, and the wave operators $\Omega_\pm(H,H_0)$ exist and are complete. In this case $\mathscr{C} \subset [-\frac{1}{4}\|\mu\|^2, 0)$ and can only accumulate in zero. (Note that for a single δ potential with strength $-\lambda$, so that μ is concentrated in a single point, $\mu < 0$ and $\|\mu\| = \lambda$, H possesses a single eigenvalue $E = \frac{1}{4}\lambda^2 = -\frac{1}{4}\|\mu\|^2$.)

Proposition 3.3: Suppose μ is real so that H is self-adjoint. Then the singular continuous spectrum of H, $\sigma_{\mathrm{sc}}(H)$, is contained in the interval $[0,\frac{1}{4}\|\mu\|^2]$. This result follows from the lemma below and the fundamental criterion (see Ref. 10, Theorem XIII.19).

Lemma 3.4: Let $\varphi(x) \in L^2 \cap L^\infty(\mathbb{R},dx)$. Then $M(z) = \varphi[z - H]^{-1}\varphi$ satisfies

$$\|M(z)\| \leqslant \tfrac{1}{2}[\,|\sqrt{z}| - \tfrac{1}{2}\|\mu\|\,]^{-1}, \quad |\sqrt{z}| > \tfrac{1}{2}\|\mu\|. \tag{3.2}$$

Proof: We note that for $\mathrm{Im}\sqrt{z} > \frac{1}{2}\|\mu\|$ the series

$$M(z) = \sum_{n=0}^{\infty} M_n(z),$$

$$M_n(z) = \varphi\,[\sqrt{z} + p]^{-1}K(z)^n[\sqrt{z} - p]^1\varphi$$

is norm convergent. On the other hand, $M_n(z)$ possesses the integral kernel

$$\langle x|M_n(z)|x'\rangle = \varphi(x)\varphi(x')\int d\mu(y_1)\cdots\int d\mu(y_n)(2i\sqrt{z})^{-1} \times\exp[i\sqrt{z}|x-y_1|]\cdots(2i\sqrt{z})^{-1} \times\exp[i\sqrt{z}|y_n - x'|],$$

so that

$$|\langle x|M_n(z)|x'\rangle| \leqslant |\varphi(x)|\varphi(x')|(2|\sqrt{z}|)^{-1}[\,\|\mu\|/(2|\sqrt{z}|)\,]^n.$$

Consequently $M_n(z)$ is Hilbert–Schmidt with Schmidt norm

$$\|M_n(z)\|_2 \leqslant (2|\sqrt{z}|)^{-1}[\,\|\mu\|/(2|\sqrt{z}|)\,]^n\|\varphi\|^2,$$

so that the series for $M(z)$ converges in Schmidt norm and hence in operator norm for $|\sqrt{z}| > \frac{1}{2}\|\mu\|$. □

Remark: If μ is absolutely continuous $d\mu(x) = V(x)dx$, with $V \in L^1(\mathbb{R}, dx)$, we can proceed as for the Rollnik class in three dimensions. In the latter case, $\|V^{1/2}[z - H_0]^{-1}V^{1/2}\|$ can be smaller than 1, uniformly in z, leading to an empty singular continuous spectrum. Here the situation is different as a result of the factor $(\sqrt{z})^{-1}$ in the integral kernel for $[z - H_0]^{-1}$; we encounter a dependence on the dimension. On the other hand, the Agmon–Kato–Kuroda theorem does have its counterpart in the present case.

Proposition 3.5: Let ν be a real regular Borel measure with finite total measure and let

$$d\mu(x) = (1 + x^2)^{-1/2-\varepsilon}d\nu(x), \quad \varepsilon > 0.$$

Then H, associated with μ, has empty singular continuous spectrum.

We omit the (lengthy) proof, which is an adaptation to the form case of the proof for the operator situation (see Ref. 10, p. 169, Theorem XIII.33, and p. 373, exercise 71). There is a further corollary to Proposition 3.1 that, although mathematically quite trivial, is important in physical applications. It allows us to perturb H with, for instance, a step potential, giving rise to interesting quantum tunneling situations, while maintaining the existence and completeness of wave operators.

Corollary 3.6: Let V be an $(H_0, 1-\varepsilon)$-bounded and $(H, 1-\varepsilon)$-bounded operator. Then $H_1 = H_0 + V$, $\mathscr{D}(H_1) = \mathscr{D}(H_0)$, $H_2 = H + V$, $\mathscr{D}(H_2) = \mathscr{D}(H)$, are closed operators and, for $z \in \rho(H_1) \cap \rho(H_2)$, $[z - H_2]^{-1} - [z - H_1]^{-1}$ is trace class, so that $\sigma_{\mathrm{ess}}(H_1) = \sigma_{\mathrm{ess}}(H_2)$. In case μ is real and V symmetric, H_1 and H_2 are self-adjoint, their absolutely continuous parts unitarily equivalent and the *generalized* [H_1 can possess further spectrum in addition to $\sigma_{\mathrm{ac}}(H_1)$] wave operators $\Omega_\pm(H_2, H_1)$ exist and are complete.

These statements directly follow from the relation, valid for suitable z,

$$[z-H_2]^{-1}-[z-H_1]^{-1}$$
$$=[1-[z-H_0]^{-1}V]^{-1}$$
$$\times\{[z-H]^{-1}-[z-H_0]^{-1}\}$$
$$\cdot[1-V[z-H]^{-1}]^{-1}. \qquad (3.3)$$

IV. A UNICITY THEOREM FOR EIGENFUNCTIONS OF H

Consider the case that $H=p^2+V(x)$, with $V(x)$ a smooth bounded function of x. The eigenvalue equation

$$Hf=Ef \qquad (4.1)$$

can be written as

$$\partial_x\begin{pmatrix}f_1\\ f_2\end{pmatrix}=\begin{pmatrix}0 & 1\\ V(x)-E & 0\end{pmatrix}\begin{pmatrix}f_1\\ f_2\end{pmatrix} \text{ or } \partial_x\mathbf{f}=\mathsf{M}(x)\cdot\mathbf{f}, \qquad (4.2)$$

where $f_1=f$, $f_2=\partial_x f$. We can interpret (4.2) as an evolution problem with the unique solution

$$\mathbf{f}(x)=\mathsf{U}(x,x_0)\cdot\mathbf{f}(x_0). \qquad (4.3)$$

Thus $f(x_0)$ and $f'(x_0)$ determine $f(x)$ and $f'(x)$, for all x; this property is referred to as unicity. In this section we shall derive a similar result for our class of form perturbations. We have to proceed in a different way, however, since we can no longer write H as an operator sum. Thus let H be associated with $\mu\in\mathfrak{M}$ as before and suppose $f\in\mathscr{D}(H)$ obeys (4.1) for some $E\in\mathbb{C}$. Then, $\forall g\in\mathscr{D}$,

$$(Hf,g)=E(f,g)=(pf,pg)+\int d\mu(x)f(x)\bar g(x)$$

[in this section $f(x)$, $g(x)$, etc., will always be the continuous version] or, with $d\nu(x)=d\mu(x)-E\,dx$,

$$-(pf,pg)=\int d\nu(x)f(x)\bar g(x),$$

so that

$$-(f,p^2g)=\int d\nu(x)f(x)\bar g(x),\quad \forall g\in\mathscr{D}_0\subset\mathscr{D}. \qquad (4.4)$$

Next we consider a special set of functions $g\in\mathscr{D}_0$, vanishing outside the finite interval (a,b), defined as follows: let $h(x)$ be a smooth function of x with the properties

$$\int_a^b dx\,x^k h(x)=0,\quad k=0,1. \qquad (4.5)$$

Thus in $\mathscr{K}=L^2[a,b],dx)$, $h\in Q\mathscr{K}$, the orthogonal complement of the subspace $P\mathscr{K}$, spanned by the constant and linear function. Now let

$$g(x)=\chi_{[a,b]}(x)\int_a^x dy(x-y)h(y)$$
$$=-\chi_{[a,b]}(x)\int_x^b dy(x-y)h(y). \qquad (4.6)$$

Then

$$g'(x)=\chi_{[a,b]}(x)\int_a^x dy\,h(y)$$
$$=-\chi_{[a,b]}(x)\int_x^b dy\,h(y),$$

and

$$g''(x)=0,\quad x\notin[a,b],\quad g''(x)=h(x),\quad x\in(a,b).$$

Although $g''(x)$ can have discontinuities in a and b, it is contained in $\mathscr{D}_0$ and

$$-(f,p^2g)=\int_a^b dx\,f(x)\,\overline{h(x)}.$$

Next we calculate the right-hand side of (4.4):

$$\int d\nu(x)f(x)\bar g(x)=-\int d\nu(x)f(x)\chi_{[a,b]}(x)\int_x^b dy(x-y)\,\overline{h(y)}$$
$$=-\int d\nu(x)f(x)\chi_{[a,b]}(x)\int_a^b dy\,\chi_{[x,b]}(y)(x-y)\,\overline{h(y)}$$
$$=-\int_a^b dy\int d\nu(x)\chi_{[a,b]}(x)f(x)\chi_{[x,b]}(y)(x-y)\,\overline{h(y)}$$
$$=\int_a^b dx\int d\nu(y)\chi_{[a,b]}(y)\chi_{[y,b]}(x)(x-y)\,\overline{h(x)}$$
$$=\int_a^b dx\int d\nu(y)\chi_{[a,x]}(y)(x-y)f(y)\,\overline{h(x)}.$$

Hence

$$\int_a^b dx\,f(x)\,\overline{h(x)}$$
$$=\int_a^b dx\int d\nu(y)\chi_{[a,x]}(y)(x-y)f(y)\bar h(x). \qquad (4.7)$$

Lemma 4.1: For $\varphi\in\mathscr{Y}=C([a,b])$, equipped with the sup-norm, define

$$(L\varphi)(x)=\int_a^x du\int d\nu(y)\chi_{[a,u)}(y)\varphi(y). \qquad (4.8)$$

Then L is a compact operator on $\mathscr{Y}$ and

$$(L\varphi)(x) = \int d\nu(y)\chi_{[a,x)}(y)(x-y)\varphi(y). \qquad (4.9)$$

Proof: Let $\|\varphi\|$ be the sup-norm of $\varphi \in \mathscr{Y}$ and let $\|\nu\| = |\nu|([a,b])$. We have

$$|(L\varphi)(x_1) - (L\varphi)(x_2)| = \left| \int_{x_1}^{x_2} du \int d\nu(y)\chi_{[a,u)}(y)\varphi(y) \right| \leqslant |x_1 - x_2| \, \|\nu\| \, \|\varphi\|,$$

so that $(L\varphi)(x)$ is continuous on $[a,b]$ and in the same way we find that $\|L\| \leqslant (b-a)\|\nu\|$. Now let $\{\varphi_n\}_{n=1}^{\infty} \subset \mathscr{Y}$ with $\|\varphi_n\| \leqslant M$. The estimate above shows that $(L\varphi_n)(x)$ is equicontinuous on $[a,b]$ and we also have $|L\varphi_n)(x)| \leqslant (b-a)\|\nu\| M$, so that, according to Ascoli's theorem, $\{L\varphi_n\}$ possesses a convergent subsequence. This settles the compactness property. Finally

$$\int_a^x du \int d\nu(y)\chi_{[a,u)}(y)\varphi(y) = \int_a^x du \int d\nu(y)\chi_{[a,x)}(y)\chi_{[a,u)}(y)\varphi(y) = \int d\nu(y)\chi_{[a,x)}(y) \int_a^x du\, \chi_{[a,u)}(y)\varphi(y) = \int d\nu(y)\chi_{[a,x)}(y)(x-y)\varphi(y),$$

which is (4.9). □

Now (4.7) takes the form

$$\int_a^b dx\, f(x)\hbar(x) = \int_a^b dx (Lf)(x)\, \overline{h(x)},$$

so that $(Pf)(x) = (PLf)(x)$, for almost every x, and hence, $f(x)$ and $(Lf)(x)$ being continuous, for every x. Consequently

$$f(x) = c + d(x-a) + (Lf)(x). \qquad (4.10)$$

Since $(Lf)(a)$ vanishes, $c = f(a)$. According to Theorem 2.4, $f(x)$ is differentiable in each $x \in \mathbb{R}\setminus C(\mu^{\mathrm{at}})$ and a glance at (4.8) shows that

$$f'(x) = d + \int d\nu(y)\chi_{[a,x)}(y)f(y), \qquad (4.11)$$

for such x. Now if $a \in \mathbb{R}\setminus C(\mu^{\mathrm{at}})$ also, then the second term in (4.11) vanishes for $x = a$. Thus $d = f'(a)$ and

$$f(x) = f(a) + f'(a)(x-a) + (Lf)(x) = f_0(x) + (Lf)(x), \quad a \in \mathbb{R}\setminus C(\mu^{\mathrm{at}}). \qquad (4.12)$$

Since L is compact, (4.12) implies

$$f(x) = ([1-L]^{-1}f_0)(x), \qquad (4.13)$$

provided that 1 is not an eigenvalue of L. Then unicity follows.

Theorem 4.2 (unicity theorem): Suppose that $a \in \mathbb{R}\setminus C(\mu^{\mathrm{at}})$ and suppose further that f is an eigenfunction of H, $Hf = Ef$. Then $f(x)$ is uniquely determined by $f(a)$ and $f'(a)$.

Proof: Let, for $x \in [a,b]$,

$$f_n(x) = (L^n f_0)(x) = \int d\nu(x_1)\chi_{[a,x)}(x_1)(x-x_1)\cdots\int d\nu(x_n)\chi_{[a,x_{n-1})}(x_n)(x_{n-1}-x_n)f_0(x_n)$$
$$= \int d\nu(x_1)\cdots\int d\nu(x_n)\theta(x-x_1)(x-x_1)\cdots\theta(x_{n-1}-x_n)(x_{n-1}-x_n)f_0(x_n),$$

where the integration intervals are $[a,x)$ (or $[a,x]$, due to the factors $(x_{j-1} - x_j)$). Next we show, by an induction argument, that

$$\max_{x>x_1\cdots>x_n>a} (x-x_1)(x_1-x_2)\cdots(x_{n-1}-x_n) \leqslant \max_{x>x_1\cdots>x_{n-1}>a} (x-x_1)(x_1-x_2)\cdots(x_{n-1}-a) \leqslant ((x-a)/n)^n \leqslant (x-a)^n/n!. \qquad (4.14)$$

We observe that

$$(x-x_1)(x_1-a) = ((x-a)/2)^2 - (x_1 - \tfrac{1}{2}(x+a))^2 \leqslant ((x-a)/2)^2.$$

Assuming that

$$(x-x_1)\cdots(x_{n-2}-x_{n-1})(x_{n-1}-a) \leqslant ((x-a)/n)^n,$$

it follows that

$$(x-x_1)\cdots(x_{n-1}-x_n)(x_n-a) \leqslant (x-x_1)((x_1-a)/n)^n \leqslant ((x-a)/(n+1))^{n+1},$$

where the last inequality follows by observing that the middle term attains its maximum for $x_1 = (nx+a)/(n+1)$. Thus (4.14) holds and consequently $[\|\nu\| = |\nu|([a,b])$, $\|\varphi\|$ is the sup-norm of $\varphi \in C([a,b])]$

$$|f_n(x)| \leqslant \|\nu\|^n (x-a)^n \|f_0\|/(n!), \qquad (4.15)$$

so the series

$$f(x) = \sum_{n=0}^{\infty} f_n(x) \qquad (4.16)$$

converges uniformly. Thus $f(x)$ is continuous and, moreover, since $f_{n+1}(x) = (Lf_n)(x)$, $(Lf)(x) = f(x) - f_0(x)$, which is (4.12). Second, (4.15) gives $\|L^n\| \leqslant c^n/(n!)$, $c = \|\nu\|(b-a)$. Thus

$$\lim_{n\to\infty} \|L^n\|^{1/n} \leqslant \lim_{n\to\infty} c(n!)^{-1/n} = 0,$$

and consequently $\sigma(L) = \{0\}$ by the spectral radius formula. Thus 1 is not an eigenvalue of L and f is given by (4.13), or, alternatively, by the expansion (4.16). □

Remark: In the preceding we only proved statements about $f(x)$ in terms of $f(a)$ and $f'(a)$, for $x \geqslant a$. The case

$x \leqslant a$ is handled in precisely the same way. Second, it is not necessary to assume $a \in \mathbb{R} \setminus C(\mu^{at})$. We can allow $a \in C(\mu^{at})$ provided we take for $f'(a)$ the right derivative in $x = a$ in (4.12).

Corollary 4.3: Under the conditions of Theorem 4.2, there exists an expression of the type (4.3) for $x \geqslant x_0 = a$ and $f_2(x)$ the right derivative of $f(x)$ [which coincides with $f'(x)$, for $x \in \mathbb{R} \setminus C(\mu^{at})$]. A similar result holds for $x \leqslant a$.

Proof: Let $u_0(x) = 1$, $u_1(x) = x$, $\psi_0(x) = ([1-L]^{-1}u_0)(x)$, and $\psi_1(x) = ([1-L]^{-1}u_1)(x)$. Then, from (4.13), for $x > a$,

$$f(x) = \psi_0(x) f(a) + \{\psi_1(x) - a\psi_0(x)\} f'(a).$$

Second, from (4.11), for $x \in \mathbb{R} \setminus C(\mu^{at}) \cap (a, \infty)$,

$$f'(x) = f'(a) + \int d\nu(y) \chi_{[a,x)}(y) \times \{\psi_0(y) f(a) + [\psi_1(y) - a\psi_0(y)] f'(a)\},$$

whose relation remains true in points of $C(\mu^{at})$ if $f'(x)$ is replaced by the right derivative. From these expressions the components of $\mathsf{U}(x)$ can be pieced together. Note that, since $(L\varphi)(a) = 0$, so that $([1-L]^{-1}\varphi)(a) = \varphi(a)$, $\mathsf{U}(a,a)$ is the unit matrix.

Corollary 4.4: Suppose that μ is concentrated in a finite or semi-infinite interval. Then $|\mathrm{Re}\sqrt{-E}| > 0$ so that in the self-adjoint case E must be strictly negative.

Proof: Without loss of generality we can suppose that $C(\mu^{at}) \subset (-\infty, 0)$. Then $d\nu(x) = -E\,dx$, for $x \geqslant 0$, and, for $x \geqslant 0$,

$$f(x) = f(0) + f'(0)(x-a) - E\int_0^x dy (x-y) f(y),$$

so that $\partial_x^2 f(x) = -Ef(x)$, for $x > 0$. Square integrability now requires $|\mathrm{Re}\sqrt{-E}| > 0$. □

Application 4.5: Let $\nu \in \mathfrak{M}$ be real with $C(\nu) \subset (-\infty, 0)$ and let $d\mu(x) = d\nu(x) - \alpha\theta(x)dx$, $\alpha > 0$. Then the self-adjoint operator H, associated with μ, has no eigenvalues $\geqslant -\alpha$.

V. DISCUSSION

The advantage of our approach through form perturbations in terms of measures is the unified treatment of both L^1- and δ-function perturbations. Although L^1 perturbations can be handled by means of a factorization procedure $[V(x) = V_1(x)V_2(x),\ V_j(x) \in L^2(\mathbb{R}, dx)]$ in a quite satisfactory way,[9] this is clearly not possible for δ potentials. The method loses its flavor if one attempts to generalize it to higher dimensions since the relation $\mathscr{D}(p) \subset C_0(\mathbb{R}^n)$ breaks down for $n > 1$. In fact, singular point interactions, such as the Fermi (or zero range) potential in three dimensions, cannot be handled in this way (a treatment for such cases in terms of resolvent expressions is given in Ref. 4). The relative ease with which we obtained a unicity (or unique continuation) theorem is also deceptive. In higher dimensions the situation is much more complicated (cf. Ref. 10, Chap. XIII and notes). In Sec. IV we mentioned the case of a measure with support in $(-\infty, 0)$ and a negative step $-\alpha\theta(x)$, $\alpha > 0$, in the origin. This is a standard example of a quantum mechanical tunneling situation. Without the step there can be an eigenvalue $E \in (-\alpha, 0)$, but, as we have seen, with the step added, this is not possible. The situation was analyzed by Howland[11] in terms of spectral concentration for the case of an L^1 potential. In the case of a δ potential with suitably chosen negative strength, centered in some $x_0 < 0$, the associated resonance pole in the resolvent can easily be calculated.[12] For sufficiently negative x_0, there exists indeed a pole $E(x_0)$ with nonzero imaginary part but as x_0 tends to zero it changes into a real number in a different Riemann sheet (a so-called virtual state). This suggest a dilatation-analytic treatment of this case (some further brief remarks can be found in Ref. 13). Here the idea is to dilate for $x \geqslant 0$ but not for $x < 0$. Then both μ and $-\alpha\theta(x)$ are not affected by the dilatation transformation but H_0 is (leading to additional boundary conditions in zero).

ACKNOWLEDGMENTS

Useful conversations with Ph. Clément, O. Diekman, H. Heymans, and P. Hofstee are gratefully acknowledged. This work is part of the research program of the Stichting voor Fundamenteel Onderzoek der Materie (Foundation for Fundamental Research on Matter) and was made possible by financial support from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Netherlands Organization for the Advancement of Research).

[1]M. Schechter, *Operator Methods in Quantum Mechanics* (North-Holland, New York, 1981).
[2]T. Kato, *Perturbation Theory for Linear Operators*, second corrected printing of the second edition (Springer, Berlin, 1984).
[3]M. Reed and B. Simon, *Methods of Modern Mathematical Physics II, Fourier Analysis, Self-adjointness* (Academic, New York, 1975).
[4]A. Grossmann, R. Høegh-Krohn, and M. Mebkhout, J. Math. Phys. **21**, 2376 (1980).
[5]B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, NJ, 1971).
[6]M. Schechter, Indiana Univ. Math. J. **22**, 483 (1972).
[7]I. W. Herbst and A. D. Sloan, Trans. Am. Math. Soc. **236**, 325 (1978).
[8]W. Rudin, *Real and Complex Analysis* (McGraw-Hill, New York, 1974).
[9]M. Reed and B. Simon, *Methods of Modern Mathematical Physics I, Functional Analysis* (Academic, New York, 1972).
[10]M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV, Analysis of Operators* (Academic, New York, 1978).
[11]J. S. Howland, Trans. Am. Math. Soc. **162**, 141 (1971).
[12]P. J. Hofstee, H. G. Muller, and A. Tip, J. Phys. A: Math. Gen. **22**, 2667 (1985).
[13]P. J. Hofstee and A. Tip, J. Math. Phys. **29**, 817 (1988).

Noncommutative differential geometry of matrix algebras

Michel Dubois-Violette
Laboratoire de Physique Théorique et Hautes Energies (Laboratoire associé au CNRS), Université de Paris-Sud, Bâtiment 211, F-91405 Orsay Cedex, France
Richard Kerner
Laboratoire de Physique Théorique des Particules Elementaires, Université Pierre et Marie Curie, 4, place Jussieu, F-75252 Paris Cedex 05, France
John Madore
Laboratoire de Physique Théorique et Hautes Energies (Laboratorie associé au CNRS), Université de Paris-Sud, Bâtiment 211, F-91405 Orsay Cedex, France

(Received 9 March 1989; accepted for publication 13 September 1989)

The noncommutative differential geometry of the algebra $M_n(\mathbb{C})$ of complex $n\times n$ matrices is investigated. The role of the algebra of differential forms is played by the graded differential algebra $C(\mathrm{sl}(n,\mathbb{C}),M_n(\mathbb{C})) = M_n(\mathbb{C}) \otimes \wedge \mathrm{sl}(n,\mathbb{C})^*$,$\mathrm{sl}(n,\mathbb{C})$ acting by inner derivations on $M_n(\mathbb{C})$. A canonical symplectic structure is exhibited for $M_n(\mathbb{C})$ for which the Poisson bracket is, to within a factor i, the commutator. Also, a canonical Riemannian structure is described for $M_n(\mathbb{C})$. Finally, the analog of the Maxwell potential is constructed and it is pointed out that there is a potential with a vanishing curvature that is not a pure gauge.

I. INTRODUCTION AND PRELIMINARIES

Let V be a smooth manifold. The differential geometry of V can be described by using the algebra $C^\infty(V)$ of smooth complex functions on V considered as an abstract commutative*-algebra. The Lie algebra of complex vector fields coincides with the Lie algebra $\mathrm{Der}(C^\infty(V))$ of derivations of $C^\infty(V)$. By definition, the Lie algebra $\mathrm{Der}(C^\infty(V))$ acts by derivations on $C^\infty(V)$, therefore[1] the complex $C(\mathrm{Der}(C^\infty(V)),C^\infty(V))$ of cochains of $\mathrm{Der}(C^\infty(V))$ with values in $C^\infty(V)$ is a graded differential algebra and one observes[2] that the graded differential algebra $\Omega(V)$ of differential forms on V is the smallest differential subalgebra of $C(\mathrm{Der}(C^\infty(V)),C^\infty(V))$ which contains $C^\infty(V)$. This led one of the authors[2] to propose the following noncommutative generalization of the differential calculus. Let $\mathscr{A}$ be an associative algebra with unit, then the complex $C(\mathrm{Der}(\mathscr{A}),\mathscr{A})$, of $\mathscr{A}$-valued cochains of the Lie algebra $\mathrm{Der}(\mathscr{A})$ of derivations of $\mathscr{A}$ is again a graded differential algebra and the smallest differential subalgebra $\Omega_D(\mathscr{A})$ of $C(\mathrm{Der}(\mathscr{A}),\mathscr{A})$ which contains $\mathscr{A}$ is a natural generalization of the algebra of differential forms, with $\mathrm{Der}(\mathscr{A})$ playing the role of the Lie algebra of vector fields.

Since there are several noncommutative generalizations of the de Rham complex,[3,4,2] it is interesting to consider the simple case where $\mathscr{A}$ is the algebra $M_n(\mathbb{C})$ of complex $n\times n$ matrices ($n\geqslant 2$). It is the aim of this paper to develop various concepts of differential geometry of $M_n(\mathbb{C})$ using $\Omega_D(M_n(\mathbb{C}))$ as algebra of differential forms. We shall show, in particular, that there is a canonical invariant symplectic form $\omega\in\Omega_D^2(M_n(\mathbb{C}))$ for which the corresponding Poisson bracket $\{\cdot,\cdot\}$ is given by $\{A,B\} = i[A,B]$. We have then, in this simple case, a precise meaning for the statement that quantum mechanics is noncommutative symplectic geometry. We shall introduce a canonical invariant Riemannian structure for $M_n(\mathbb{C})$, describe the corresponding Hodge theory on $\Omega_D(M_n(\mathbb{C}))$ and, in the case $n=2$, diagonalize the Laplacian. Finally, we shall also describe the analog of Maxwell's electromagnetic potential: It is a Hermitian connection on the free Hermitian $M_n(\mathbb{C})$ module of rank one. We show that, in contrast to the commutative case, there is a unique potential with vanishing curvature which is not a pure gauge. This potential is gauge invariant and is related to the above-mentioned canonical invariant symplectic form.

The plan of the paper is as follows. In Sec. II, we describe in some generality, the differential calculus that we use. In Sec. III, we give a presentation in terms of generators and relations of our analog of the differential algebra of differential forms. In Sec. IV, we describe the canonical symplectic structure of $M_n(\mathbb{C})$. In Sec. V, the analog of integration theory is introduced. Section VI deals with the canonical Riemannian structure and the corresponding Hodge–de Rham theory. In Sec. VII, we construct the analog of electromagnetism. In Sec. VIII, we diagonalize the Laplacian on forms in the case of 2×2 matrices. Section IX contains our conclusions.

Our notations are more or less standard. We use the Einstein convention of summation of repeated up–down indices.

II. DIFFERENTIAL CALCULUS FOR $M_n(\mathbb{C})$

A. The graded differential algebra $\Omega_D(M_n(\mathbb{C}))$

Any derviation of $M_n(\mathbb{C})$ is an inner derivation thus the Lie algebra $\mathrm{Der}(M_n(\mathbb{C}))$ identifies canonically with $\mathrm{sl}(n,\mathbb{C})$. It was pointed out in Ref. 2 that the smallest differential subalgebra $\Omega_D(M_n(\mathbb{C}))$ of $C(\mathrm{Der}(M_n(\mathbb{C})), M_n(\mathbb{C}))$ which contains $M_n(\mathbb{C})$ is $C(\mathrm{Der}(M_n(\mathbb{C})), M_n(\mathbb{C}))$ itself. Therefore, one has $\Omega_D(M_n(\mathbb{C})) = C(\mathrm{Der}(M_n(\mathbb{C})) = M_n(\mathbb{C}) \otimes \wedge \mathrm{sl}(n,\mathbb{C})^*$. An element α of $\Omega_D^p(M_n(\mathbb{C}))$ is a p-linear antisymmetric mapping of $\mathrm{Der}(M_n(\mathbb{C}))$ to $M_n(\mathbb{C})$, $(X_1,\ldots,X_p)\mapsto\alpha(X_1,\ldots,X_p)\in M_n(\mathbb{C})$, and its differential $d\alpha\in\Omega_D^{p+1}(M_n(\mathbb{C}))$ is given by[5,6,1]:

 0022-2488/90/020316-07$03.00

$$d\alpha(X_0,X_1,\ldots,X_p) = \sum_{0<k<p} (-1)^k X_k \alpha(X_0,\cdots,\overset{k}{\vee},\cdots,X_p)$$

$$+ \sum_{0<r<s<p} (-1)^{r+s} \alpha([X_r,X_s],X_0,\cdots,\overset{r}{\vee}\cdots\overset{s}{\vee},\cdots,X_p)$$

for $X_0,\ldots,X_p \in \mathrm{Der}(M_n(\mathbb{C}))(=\mathrm{sl}(n,\mathbb{C}))$, (where $\overset{k}{\vee}$ means that the k th term is omitted).

B. The cohomology of $\Omega_D(M_n(\mathbb{C}))$

The only elements of $M_n(\mathbb{C})$ invariant under $\mathrm{Der}(M_n(\mathbb{C}))$ [i.e., by the adjoint action of $\mathrm{sl}(n,\mathbb{C})$] are the multiples of $\mathbf{1}\in M_n(\mathbb{C})$ thus, it follows from the semisimplicity of $\mathrm{sl}(n,\mathbb{C})$ that the cohomology $H_D(M_n(\mathbb{C}))$ of $\Omega_D(M_n(\mathbb{C}))$ identifies with the Lie algebra cohomology $H^\bullet(\mathrm{sl}(n,\mathbb{C}))$ of $\mathrm{sl}(n,\mathbb{C})$ (Ref. 2). This cohomology is well known (Ref. 1); it is the free graded-commutative algebra with unit $\wedge(c_3,\ldots,c_{2n-1})$ generated by elements c_{2p-1}, $p\in\{2,3,\ldots,n\}$, with c_{2p-1} of degree $2p-1$. In particular, one has

$$H^1_D(M_n(\mathbb{C})) = H^2_D(M_n(\mathbb{C})) = 0$$

and

$$H^0_D(M_n(\mathbb{C})) = H^3_D(M_n\mathbb{C})) = H^{n^2-1}_D(M_n(\mathbb{C})) = \mathbb{C}.$$

C. The operation of $\mathrm{Der}(M_n(\mathbb{C}))$ in $\Omega_D(M_n(\mathbb{C}))$

As in the general case,[2] there is an operation of the Lie algebra $\mathrm{Der}(M_n(\mathbb{C}))$ in the graded differential algebra $\Omega_D(M_n(\mathbb{C}))$ in the sense of Ref. 7 which we now describe. For any $X\in\mathrm{Der}(M_n(\mathbb{C}))$, one verifies that one defines an antiderivation i_X of degree -1 of $\Omega_D(M_n(\mathbb{C}))$ by $i_X\alpha(X_1,\ldots,X_{p-1}) = \alpha(X,X_1,\ldots,X_{p-1})$ for $\alpha\in\Omega^p_D(M_n(\mathbb{C}))$ with $p\geqslant 1, X_i\in\mathrm{Der}(M_n(\mathbb{C}))$ and $i_X\Omega^0_D(M_n(\mathbb{C}))=0$. Then, $L_X = di_X + i_X d$ is a derivation of degree 0 of $\Omega_D(M_n(\mathbb{C}))$ which extends X. Here, i_X is the analog of the inner product of forms by a vector field and L_X is the analog of the Lie derivative of forms by a vector field. One has the following characteristic relations of operations[7]:

$$i_{X_1}i_{X_2} + i_{X_2}i_{X_1} = 0,\quad [L_{X_1},i_{X_2}] = i_{[X_1,X_2]}$$

and

$$[L_{X_1},L_{X_2}] = L_{[X_1,X_2]},\quad \forall X_1,X_2\in\mathrm{Der}(M_n(\mathbb{C})).$$

An element α of $\Omega_D(M_n(\mathbb{C}))$ will be said to be *invariant* if $L_X\alpha = 0$ for any $X\in\mathrm{Der}(M_n(\mathbb{C}))$; invariant elements form a graded differential subalgebra with unit of $\Omega_D(M_n(\mathbb{C}))$.

III. PRESENTATION ASSOCIATED TO A BASIS

A. Basis for $M_n(\mathbb{C})$

Let $E_k, k\in\{1,2,\ldots,n^2-1\}$ be a basis of Hermitian traceless $n\times n$ matrices. Then $\mathbf{1}, E_1,\ldots,E_{n^2-1}$ is a basis of $M_n(\mathbb{C})$ consisting of Hermitian matrices. One has a multiplication table of the form

$$E_k E_l = g_{kl}\mathbf{1} + s^m_{kl}E_m - (i/2)\, C^m_{kl}E_m, \qquad (1)$$

where $g_{kl} = g_{lk} = (1/n)\mathrm{Tr}(E_k E_l)$, $s^m_{kl} = s^m_{lk}$, and $C^m_{kl} = -C^m_{lk}$ are real numbers. These numbers are canonically the components of three ad-invariant $\mathrm{SU}(n)$ tensors. Thus $s^n_{kn} = C^n_{kn} = 0$ and $g_{kl}, s^n_{km}s^m_{ln}, C^n_{km}C^m_{ln}$ are the components of three bilinear forms proportional to the Killing form of $\mathrm{SU}(n)$. Finally, $(s^n_{kl} - (i/2)\, C^n_{kl})g_{nm} = (1/n)\mathrm{Tr}(E_k E_l E_m)$ implies that $s_{klm} = s^n_{kl}g_{nm}$ is completely symmetric and that $C_{klm} = C^n_{kl}g_{nm}$ is completely antisymmetric.

B. Associated basis of $\mathrm{Der}(M_n(\mathbb{C}))$

Setting $\partial_k = \mathrm{ad}(iE_k)$, the ∂_k, $k\in\{1,2,\ldots,n^2-1\}$, form a basis of $\mathrm{Der}(M_n(\mathbb{C})) = \mathrm{sl}(n,\mathbb{C})$ and one has $[\partial_k,\partial_l] = C^m_{kl}\partial_m$. The real combinations of the ∂_k's form a real Lie algebra $\mathrm{Der}_{\mathbb{R}}(M_n(\mathbb{C}))$, which identifies to $\mathrm{su}(n)$; these derivations of $M_n(\mathbb{C})$ are real in the sense that one has $\partial_k(A^*) = (\partial_k A)^*$ for $A\in M_n(\mathbb{C})$, i.e., they preserve hermiticity.

C. Generators of $\Omega_D(M_n(\mathbb{C}))$

Let $\theta^k\in\Omega^1_D(M_n(\mathbb{C}))$, $k\in\{1,2,\ldots,n^2-1\}$, be defined by $\theta^k(\partial_l) = \delta^k_l\cdot\mathbf{1}$, i.e., (θ^k) is the dual basis of (∂_k) in $\mathrm{sl}(n,\mathbb{C})^*$ identified to $\mathbf{1}\otimes\mathrm{sl}(n,\mathbb{C})^*\subset\Omega^1_D(M_n(\mathbb{C}))$. One has in $\Omega_D(M_n(\mathbb{C}))$.

$$E_k\theta^l = \theta^l E_k \qquad (2)$$

and

$$\theta^k\theta^l = -\theta^l\theta^k. \qquad (3)$$

The differential d of $\Omega_D(M_n(\mathbb{C}))$ is then given by

$$dE_k = -C^m_{kl}E_m\theta^l \qquad (4)$$

and

$$d\theta^k = -\tfrac{1}{2}C^k_{lm}\theta^l\theta^m \qquad (5)$$

($d^2=0$ follows from the Jacobi identity). Relations (1) to (5) for generators E_k, θ^l and differential d give a presentation of $\Omega_D(M_n(\mathbb{C}))$.

It is worth noticing here that one could see the dE_k's instead of the θ^k's as generators of $\Omega^1_D(M_n(\mathbb{C}))$ as left (or right) $M_n(\mathbb{C})$ module; however their commutation properties in $\Omega_D(M_n(\mathbb{C}))$ are more complicated, in particular, $E_k dE_l \neq (dE_l)E_k$.

D. Reality

When $M_n(\mathbb{C})$ is the analog of complex functions, the analog of a real function is the real subspace of Hermitian matrices. Thus the analog of real vector fields are the real derivations in the sense of Sec. III B and the θ^k's must be considered as real. Therefore, one is led to define an antilinear involutive mapping of $\Omega_D(M_n(\mathbb{C}))$ on itself, $\alpha\mapsto\bar\alpha$, by

$$\overline{A\theta^{i_1}\cdots\theta^{i_p}} = A^*\theta^{i_1}\cdots\theta^{i_p}.$$

The elements α of $\Omega_D(M_n(\mathbb{C}))$ satisfying $\alpha=\bar\alpha$ will be called real elements. The real vector space of real elements of $\Omega_D(M_n(\mathbb{C}))$ is the analog of the space of real differential forms, [$\Omega_D(M_n(\mathbb{C}))$ being the analog of the algebra of complex differential forms]. Notice that the definition of $\alpha \mapsto \overline{\alpha}$ is independent of the chosen basis (E_k) of Hermitian traceless matrices; furthermore, if α is real, then $d\alpha$ is also real $(d\overline{\alpha} = \overline{d\alpha})$.

IV. CANONICAL SYMPLECTIC STRUCTURE

A. The canonical invariant element θ of $\Omega^1_D(M_n(\mathbb{C}))$

With the notations of Sec. III, consider the real element $\theta = E_k\theta^k$ of $\Omega^1_D(M_n(\mathbb{C}))$. The first observation is that θ is independent from the chosen basis (E_k) of Hermitian traceless matrices. The second almost obvious observation is that θ is invariant. Since we shall need at several places the expression of $d\theta$, let us check carefully this last point.

One has $i_{\partial_k}\theta = E_k$ so, in view of (4), $di_{\partial_k}\theta = -C^m_{kl}E_m\theta^l$. On the other hand one has, in view of (4), (5):

$$d\theta = \tfrac{1}{2}C^k_{lm}E_k\theta^l\theta^m \tag{6}$$

or

$$d\theta = i(\theta)^2. \tag{6'}$$

Thus $i_{\partial_k}d\theta = C^m_{kl}E_m\theta^l$ and therefore $L_{\partial_k}\theta = (di_{\partial_k} + i_{\partial_k}d)\theta = 0$, which shows that θ is invariant. Moreover, any invariant element of $\Omega^1_D(M_n(\mathbb{C}))$ is a (complex) multiple of θ.

B. Symplectic structures for $M_n(\mathbb{C})$

Let V be a smooth manifold and recall that a symplectic structure (or a symplectic form) on V is a real closed nondegenerated two-form ω. Given such a two-form, one defines the Hamiltonian vector field $\mathrm{Ham}(f)$ associated to $f \in C^\infty(V)$ by $\omega(X,\mathrm{Ham}(f)) = Xf$, for any vector field X, and one defines the Poisson bracket $\{f,g\}$ of $f,g \in C^\infty(V)$ by $\{f,g\} = \omega(\mathrm{Ham}(f),\mathrm{Ham}(g))$; $d\omega = 0$ implies the Jacobi identity for $\{\cdot,\cdot\}$.

Here, the rules of the game are that $M_n(\mathbb{C})$ is the analog of the algebra of smooth functions, $\mathrm{Der}(M_n(\mathbb{C}))$ is the analog of the Lie algebra of vector fields and $\Omega_D(M_n(\mathbb{C}))$ is the analog of the algebra of differential forms. Therefore, it is natural to call symplectic structure a real closed element ω of $\Omega^2_D(M_n(\mathbb{C}))$ such that for each $A \in M_n(\mathbb{C})$, $\omega(X,\mathrm{Ham}(A)) = XA$ $\forall X \in \mathrm{Der}(M_n(\mathbb{C}))$ has a unique solution $\mathrm{Ham}(A) \in \mathrm{Der}(M_n(\mathbb{C}))$. One defines the Poisson bracket $\{A,B\}$ of $A,B \in M_n(\mathbb{C})$ by $\{A,B\} = \omega(\mathrm{Ham}(A),\mathrm{Ham}(B))$. One has, by definition $\{A,B\} = \mathrm{Ham}(A)B = -\mathrm{Ham}(B)A$; using this, one sees that $d\omega(\mathrm{Ham}(A),\mathrm{Ham}(B),\mathrm{Ham}(C)) = 0$ is the Jacobi identity $\{A,\{B,C\}\} + \{B,\{C,A\}\} + \{C,\{A,B\}\} = 0$ and that $d\omega(X,\mathrm{Ham}(A),\mathrm{Ham}(B)) = 0$ $\forall X \in \mathrm{Der}(M_n(\mathbb{C}))$ is equivalent to $[\mathrm{Ham}(A),\mathrm{Ham}(B)] = \mathrm{Ham}(\{A,B\})$. Furthermore, it also follows from $\{A,B\} = \mathrm{Ham}(A)B$ that one has $\{A,BC\} = \{A,B\}C + B\{A,C\}$. Everything works as in the commutative case.

Notice that ω must be exact since $H^2_D(M_n(\mathbb{C})) = 0$.

C. The canonical symplectic structure

It is natural to consider the closed real invariant element $\omega = d\theta$ of $\Omega^2_D(M_n(\mathbb{C}))$. From formula (6), it follows that one has $\omega(\partial_k,\partial_l) = C^m_{kl}E_m = i[E_k,E_l] = \partial_k E_l$. This shows that ω satisfies the conditions of Sec. IV B and that $\mathrm{Ham}(A) = \mathrm{ad}(iA)$, $A \in M_n(\mathbb{C})$. Thus the corresponding Poisson bracket $\{\cdot,\cdot\}$ is given by

$$\{A,B\} = \omega(\mathrm{Ham}(A),\mathrm{Ham}(B)) = i[A,B], \quad A,B \in M_n(\mathbb{C}). \tag{7}$$

We refer to this structure as *the canonical symplectic structure* for $M_n(\mathbb{C})$.

D. Remark

If one replaces θ by $\hbar\theta$ for some $\hbar \in \mathbb{R}$, i.e., ω by $\hbar\omega$, then $\{A,B\} = i[A,B]$ is replaced by $\{A,B\} = (i/\hbar)[A,B]$.

V. INTEGRATION: THE CYCLE $(\Omega_D(M_n(\mathbb{C})),\int)$

A. Structure of $\Omega^{n^2-1}_D(M_n(\mathbb{C}))$.

As left (or right) $M_n(\mathbb{C})$ module, $\Omega^{n^2-1}_D(M_n(\mathbb{C}))$ is generated by $\theta^1\theta^2\cdots\theta^{n^2-1}$, however, this element depends on the basis (E_k). Let $g = \det(g_{kl})$ be the determinant of the real positive definite $(n^2-1)\times(n^2-1)$ matrix defined in Eq. (1). Then $\sqrt{g}\theta^1\cdots\theta^{n^2-1} \in \Omega^{n^2-1}_D(M_n(\mathbb{C}))$ does only depend on the orientation of the basis (E_k); thus this real element is intrinsically defined up to a factor ± 1 and one fixes it by choosing an orientation. An arbitrary element $\alpha \in \Omega^{n^2-1}_D(M_n(\mathbb{C}))$ is of the form $A\sqrt{g}\,\theta^1\cdots\theta^{n^2-1}$ for a unique $A \in M_n(\mathbb{C})$; we define a linear mapping $\int : \Omega^{n^2-1}_D(M_n(\mathbb{C})) \to \mathbb{C}$ by $\int\alpha = (1/n)\mathrm{Tr}(A)$, where α and A are as above.

Lemma: (a) $\int$ is a closed graded trace, i.e., one has

$$\int d\beta = 0, \quad \forall \beta \in \Omega^{n^2-2}_D(M_n(\mathbb{C}))$$

and

$$\int \sigma\tau = (-1)^{pq}\int \tau\sigma$$

for

$\sigma \in \Omega^p_D(M_n(\mathbb{C}))$ and $\tau \in \Omega^q_D(M_n(\mathbb{C}))$ with $p + q = n^2 - 1$.

(b) $\sqrt{g}\theta^1\theta^2\cdots\theta^{n^2-1}$ is invariant i.e., $L_X(\sqrt{g}\theta^1\theta^2\cdots\theta^{n^2-1}) = 0$, $\forall X \in \mathrm{Der}(M_n\mathbb{C}))$. Therefore, (a) means that $(\Omega_D(M_n(\mathbb{C})),\int)$ is *a cycle of dimension* n^2-1 in the sense of Ref. 3.

Proof: It follows from formula (5) and from the complete antisymmetry of $C_{klm} = C^n_{kl}g_{nm}$ that one has $d(\theta^{i_1}\cdots\theta^{i_{n^2-2}}) = 0$, this implies statement (b) of the lemma and shows that the only contributions to $d\beta$ comes from the differential of linear combinations of terms of the form $E_k(\theta^{i_1}\cdots\theta^{i_{n^2-2}})$ but then, formula (4) shows that $d\beta$ is of the form $E\theta^1\cdots\theta n^{2-1}$ for some traceless matrix E. Therefore, $\int d\beta = 0$ follows from the definition of $\int$. □

VI. CANONICAL RIEMANNIAN STRUCTURE

A. The metric

The element $\sqrt{g}\theta^1\cdots\theta^{n^2-1}$ of $\Omega_D^{n^2-1}(M_n(\mathbb{C}))$ introduced in Sec. V A looks like the volume element of a metric. This suggests to introduce the object

$$g_{kl}\theta^k\otimes\theta^l\in M_n(\mathbb{C})\otimes S^2\mathrm{sl}(n,\mathbb{C})^*$$
$$\times(\subset\Omega_D^1(M_n(\mathbb{C}))\underset{M_n(\mathbb{C})}{\otimes}\Omega_D^1(M_n(\mathbb{C}))).$$

Here, $g_{kl}\theta^k\otimes\theta^l$ is really the analog of an invariant Riemannian metric [for $M_n(\mathbb{C})$] and we shall call this structure *the canonical Riemannian structure*. The inverse matrix of (g^{kl}) will be denoted by (g^{kl}).

B. The star isomorphism

One defines a linear mapping

$$*:\Omega_D(M_n(\mathbb{C}))\to\Omega_D(M_n(\mathbb{C}))$$

by

$$*(\theta^{i_1}\cdots\theta^{i_p})=\frac{1}{(n^2-1-p)!}\sqrt{g}g^{i_1j_1}\cdots g^{i_pj_p}$$
$$\times\epsilon_{j_1,\ldots,j_{n^2-1}}\theta^{j_{p+1}}\cdots\theta^{j_{n^2-1}},$$

and by

$$*(A\theta^{i_1}\cdots\theta^{i_p})=A*(\theta^{i_1}\cdots\theta^{i_p})\quad\text{for}\quad A\in M_n(\mathbb{C})$$

(where $\epsilon_{j_1,\ldots,j_{n^2-1}}$ is completely antisymmetric with $\epsilon_{1,2,\ldots,n^2-1}=1$). One has $*(\Omega_D^p(M_n(\mathbb{C}))\subset\Omega_D^{n^2-1-p}\times(M_n(\mathbb{C})))$ and $*(*\alpha)=(-1)^{n^2p}\alpha$ if $\alpha\in\Omega_D^p(M_n(\mathbb{C}))$.

Setting $(\alpha|\beta)=\int\alpha\cdot*(\beta)$ if α and β are of the same degree and $(\alpha|\beta)=0$ otherwise, one has $(\alpha|\beta)=(\beta|\alpha)$ in view of the graded trace property of $\int$ and this inner product is a real positive-definite bilinear form on the real subspace of real elements of $\Omega_D(M_n(\mathbb{C}))$. It follows that $(\alpha,\beta)\mapsto\langle\alpha|\beta\rangle=(\bar{\alpha}|\beta)$ is a positive-definite Hermitian form on $\Omega_D(M_n(\mathbb{C}))$; so, equipped with $\langle.|.\rangle$, $\Omega_D(M_n(\mathbb{C}))$ is a (graded finite-dimensional) complex Hilbert space.

C. The Laplacian

We define $\delta:\Omega_D(M_n(\mathbb{C}))\to\Omega_D(M_n(\mathbb{C}))$ by $\langle d\alpha|\beta\rangle=\langle\alpha|\delta\beta\rangle$, $\forall\alpha,\beta\in\Omega_D(M_n(\mathbb{C}))$. One verifies by using the fact that $\int$ is closed that δ is given by

$$\delta\alpha=(-1)^{(n^2-1)p+n^2}*d*\alpha\text{ for }\alpha\in\Omega_D^p(M_n(\mathbb{C})).$$

We define the *Laplacian* Δ on $\Omega_D(M_n(\mathbb{C}))$ by $\Delta=d\delta+\delta d$. From $\langle\alpha|\Delta\alpha\rangle=\langle\delta\alpha|\delta\alpha\rangle+\langle d\alpha|d\alpha\rangle=\|\delta\alpha\|^2+\|d\alpha\|^2$, it follows that $\Delta\geqslant0$ as operator on the Hilbert space $(\Omega_D(M_n(\mathbb{C})),\langle.|.\rangle)$. It also follows that $\Delta\alpha=0$ *if and only if* $d\alpha=0$ *and* $\delta\alpha=0$; such an α will be called harmonic, the space $\mathrm{Ker}(\Delta)$ of these elements is the kernel of Δ.

By definition, the orthogonal complement of $\delta\Omega_D(M_n(\mathbb{C}))$ is the space of $\alpha\in\Omega_D(M_n(\mathbb{C}))$ satisfying $d\alpha=0$ and the orthogonal complement of $d\Omega_D(M_n(\mathbb{C}))$ is the space $\alpha\in\Omega_D(M_n(\mathbb{C}))$ satisfying $\delta\alpha=0$. It follows that *one has a decomposition* $\Omega_D(M_n(\mathbb{C}))=d\Omega_D(M_n(\mathbb{C}))\oplus\delta\Omega_D(M_n(\mathbb{C}))\oplus\mathrm{Ker}(\Delta)$ *of* $\Omega_D(M_n(\mathbb{C}))$ *in three orthogonal subspaces* (for $\langle.|.\rangle$) which is the analog of Hodge–de Rham decomposition.

Proposition: The linear mapping of $\mathrm{Ker}(\Delta)$ in $H_D(M_n(\mathbb{C}))$ which associates to $\alpha\in\mathrm{Ker}(\Delta)$ its class $[\alpha]$ in $H_D(M_n(\mathbb{C}))$ is an isomorphism of graded vector spaces. Furthermore, $\alpha\in\mathrm{Ker}(\Delta)$ if and only if α is an invariant element of the subalgebra $\mathbf{1}\otimes\wedge\mathrm{sl}(n,\mathbb{C})^*$ of $\Omega_D(M_n(\mathbb{C}))$ generated by the θ^k, $k\in\{1,2,\ldots,n^2-1\}$.

Proof: Let $\alpha\in\mathrm{Ker}(\Delta)$, then if $\alpha+d\beta\in\mathrm{Ker}(\Delta)$ one has $\delta d\beta=0$ so $\langle\beta|\delta d\beta\rangle=\|d\beta\|^2=0$ which implies $d\beta=0$. This shows that $\alpha\mapsto[\alpha]$ is injective from $\mathrm{Ker}(\Delta)$ in $H_D(M_n(\mathbb{C}))$.

The subalgebra $\mathbf{1}\otimes\wedge\mathrm{sl}(n,\mathbb{C})^*$ of $\Omega_D(M_n(\mathbb{C}))$ generated by the θ^k is a differential subalgebra of $\Omega_D(M_n(\mathbb{C}))$. Let $\mathscr{I}$ denote the algebra of the invariant elements of $\mathbf{1}\otimes\wedge\mathrm{sl}(n,\mathbb{C})^*$. By using the Koszul formula[6] and the definitions of δ, one checks that, if $\alpha\in\mathscr{I}$ one has $d\alpha=0$ and $\delta\alpha=0$. On the other hand, one knows[6] that $\alpha\mapsto[\alpha]$ is a bijection of $\mathscr{I}$ onto $H_D(M_n(\mathbb{C}))=H^\bullet(\mathrm{sl}(n,\mathbb{C}))$. This shows that $\alpha\mapsto[\alpha]$ is surjective and therefore bijective from $\mathrm{Ker}(\Delta)$ onto $H_D(M_n(\mathbb{C}))$ and that $\mathrm{Ker}(\Delta)$ therefore coincides with $\mathscr{I}$. □

Remark: The last statements have a classical geometrical interpretation. If one identifies the θ^k with the components of the Maurer–Cartan form of $SU(n)$, then $\mathbf{1}\otimes\wedge\mathrm{sl}(n,\mathbb{C})^*$ becomes identified with the differential algebra of left-invariant forms on $SU(n)$ and $\mathscr{I}$ with the algebra of bi-invariant forms on $SU(n)$. Then, $g_{kl}\theta^k\otimes\theta^k$ is up to a factor the metric of $SU(n)$ and on $SU(n)$, the harmonic forms are the bi-invariant forms (this is true for any compact semisimple Lie group).

VII. CONNECTIONS ON THE FREE HERMITIAN $M_n(\mathbb{C})$ MODULE OF RANK ONE

A. The free Hermitian module $\mathscr{H}$

Let $\mathscr{M}$ be a right $M_n(\mathbb{C})$ module. A *Hermitian structure* on $\mathscr{M}$ (Ref. 8) is an $M_n(\mathbb{C})$-valued positive definite Hermitian form $h(\phi,\psi)\in M_n(\mathbb{C})$ $\phi,\psi\in\mathscr{M}$ such that $h(\phi A,\psi B)=A^*h(\phi,\psi)B$, $\forall\phi,\psi\in\mathscr{M}$, $\forall A,B\in M_n(\mathbb{C})$. The pair $(\mathscr{M},h)$, or simply $\mathscr{M}$ if there is no ambiguity, will be called a Hermitian module. A Ω_D*-connection*, or simply a *connection* on $\mathscr{M}$ (Ref. 3), is a linear mapping

$$\nabla:\mathscr{M}\to\mathscr{M}\underset{M_n(\mathbb{C})}{\otimes}\Omega_D^1(M_n(\mathbb{C}))$$

such that $\nabla(\phi A)=(\nabla\phi)A+\phi\otimes dA$, $\forall\phi\in\mathscr{M}$, $\forall A\in M_n(\mathbb{C})$. Here, ∇ is a *Hermitian connection* on $(\mathscr{M},h)$ if $dh(\phi,\psi)=h(\nabla\phi,\psi)+h(\phi,\nabla\psi)$, $\forall\phi,\psi\in\mathscr{M}$. Connections always exist on projective modules of finite type.[3]

We denote by $\mathscr{H}$ the simplest Hermitian $M_n(\mathbb{C})$-module, namely, the free Hermitian module of rank one. An element e of $\mathscr{H}$ such that $h(e,e)=\mathbf{1}$ will be called a *unitary generator of* $\mathscr{H}$ or a *gauge*. One then has $\mathscr{H}=eM_n(\mathbb{C})$ and $h(eA,eB)=A^*B$, $\forall A,B\in M_n(\mathbb{C})$. Here, $U\mapsto eU$ is a bijection of the group $U(n)$ of unitary elements of $M_n(\mathbb{C})$ onto the set of unitary generators of $\mathscr{H}$. Such a change of unitary generator is a *gauge transformation*. The group of gauge transformations is therefore $U(n)$.

B. Connections on $\mathscr{H}$

Let ∇ be a Hermitian connection on $\mathscr{H}$. Given a gauge e, any $\phi\in\mathscr{H}$ is of the form $\phi = eB$ for a unique $B\in M_n(\mathbb{C})$; thus $\nabla\phi = (\nabla e)B + e\otimes dB$ where $\nabla e = e\otimes\alpha$ for a unique $\alpha\in\Omega^1_D(M_n(\mathbb{C}))$ satisfying $\bar\alpha = -\alpha$ in view of the Hermitian property of ∇; B and α as above will be called the components of ϕ and ∇ in e. Under a change of gauge $e\mapsto eU$, $U\in U(n)$, they transform according to the rules:

$$B\mapsto U^{-1}B \quad\text{and}\quad \alpha\mapsto U^{-1}\alpha U + U^{-1}dU.$$

Remember that $M_n(\mathbb{C})$ is the analog of $C^\infty(V)$. So $U(n)$ is the analog of $U(1)$-valued functions on V, one sees that what we are introducing is the analog of electromagnetism. Namely, B and α as above are, respectively, the analogs of the component of a charged scalar field and Maxwell potential in a given gauge.

Given a gauge e, there is a unique connection $\overset{(e)}{\nabla}$ on $\mathscr{H}$ such that $\overset{(e)}{\nabla} e = 0$ (i.e., $\overset{(e)}{\nabla}(eB) = e\otimes dB$, $\forall B\in M_n(\mathbb{C})$); the component of $\overset{(e)}{\nabla}$ in e vanishes and its component in an arbitrary gauge eU is given by $U^{-1}dU$. Conversely, if the component α of a connection ∇ on $\mathscr{H}$ in a gauge e is $\alpha = U^{-1}dU$ for some $U\in U(n)$ then one has $\nabla = \overset{(eU^{-1})}{\nabla}$. These connections $\overset{(e)}{\nabla}$ when e runs over the set of unitary generators of $\mathscr{H}$ (i.e., over the gauges) will be called pure gauge connections, they are automatically Hermitian connections on $\mathscr{H}$.

If ∇ and ∇' are connections, one has as usual $(\nabla-\nabla')(\phi B) = ((\nabla-\nabla')\phi)B$ so $\nabla-\nabla'$ is a right module homomorphism; if α and α' are the components of ∇ and ∇' in a gauge e then under a gauge transformtion $e\mapsto eU$, $\alpha-\alpha'$ transforms homogeneously as $\alpha-\alpha'\mapsto U^{-1}(\alpha-\alpha')U$. Connections on $\mathscr{H}$ form an affine space, however, here there is a natural origin $\overset{0}{\nabla}$ in this affine space which we now introduce.

Lemma: Define the linear mapping $\overset{0}{\nabla}:\mathscr{H}\to\mathscr{H}\otimes_{M_n(\mathbb{C})}\Omega^1_D(M_n(\mathbb{C}))$ by $\overset{0}{\nabla}\phi = -i\phi\otimes\theta$, $\forall\phi\in\mathscr{H}$, where θ is the canonical invariant element of $\Omega^1_D(M_n(\mathbb{C}))$ defined by Sec. IV A. Then, $\overset{0}{\nabla}$ is a Hermitian connection on $\mathscr{H}$ which is gauge invariant in the sense that $\overset{0}{\nabla} e = \overset{0}{\nabla}(eU)$, $\forall U\in U(n)$; in fact, $\overset{0}{\nabla} e = e\otimes(-i\theta)$ and $U^{-1}(-i\theta)U + U^{-1}dU = -i\theta$, $\forall U\in U(n)$. Let ∇ be a gauge invariant connection on $\mathscr{H}$ [i.e., $\nabla e = \nabla(eU)$, $\forall U\in U(n)$], then one has $\nabla\phi = -i\phi\otimes(\theta + \lambda_k\theta^k)$ for some $\lambda_k\in\mathbb{C}$, $k = 1,2,\ldots,n^2-1$; Furthermore, ∇ is Hermitian if and only if $\lambda_k\in\mathbb{R}$, $\forall k\in\{1,2,\ldots,n^2-1\}$.

Proof: In terms of $\theta = E_k\theta^k$, formula (4) can be written

$$dB = i[\theta,B],\quad \forall B\in M_n(\mathbb{C}). \tag{4'}$$

This implies $\overset{0}{\nabla}(\phi B) = -i\ (\phi B)\theta = (i\phi\theta)B + \phi i[\theta,B] = (\overset{0}{\nabla}\phi)B + \phi\, dB$, so $\overset{0}{\nabla}$ is a connection which is obviously Hermitian. Its component in any gauge is, by definition, $-i\theta$ and one verifies directly, by using (4′) that $U^{-1}(-i\theta)U + U^{-1}dU = -i\theta$.

Let ∇ be any connection on $\mathscr{H}$ and let $-i(\theta+\beta)\in\Omega^1_D(M_n(\mathbb{C}))$ be its component in gauge e. Then, its component in gauge eU, $U\in U(n)$, is $-i(\theta + U^{-1}\beta U)$ so ∇ is gauge invariant if and only if one has $\beta = U^{-1}\beta U$, $\forall U\in U(n)$, which implies $\beta = \lambda_k\theta^k$ with $\lambda_k\in\mathbb{C}$, $\forall k$. Finally, ∇ is Hermitian if and only if $\bar\beta = \beta$, i.e., $\lambda_k\in\mathbb{R}$, $\forall k$. □

Here, $\overset{0}{\nabla}$ will be called *the canonical connection on* $\mathscr{H}$.

Remarks: (a) $\overset{0}{\nabla}$ cannot be a pure gauge connection [i.e., there is no $U\in U(n)$ such that $-i\theta = U^{-1}dU$], since it is gauge invariant [i.e., one has $U^{-1}(-i\theta)U + U^{-1}dU = -i\theta$, $\forall U\in U(n)$].

(b) $\overset{0}{\nabla}$ is gauge invariant and Hermitian but is not unique under these conditions since these properties are also true if one replaces θ by $\theta + \lambda_k\theta^k$ for $\lambda_k\in\mathbb{R}$. However $\overset{0}{\nabla}$ is completely specified by the fact that it is the only connection with vanishing curvature which is not a pure gauge connection. We now describe curvature for connections on $\mathscr{H}$.

C. Curvature

Let ∇ be a connection on $\mathscr{H}$. Then, one extends ∇ as a linear mapping, again denoted by ∇, of $\mathscr{H}\otimes_{M_n(\mathbb{C})}\Omega_D(M_n(\mathbb{C}))$ in itself by setting[3]

$$\nabla(\phi\otimes\alpha) = (\nabla\phi)\alpha + \phi\otimes d\alpha,\quad \forall\phi\in\mathscr{H},\ \forall\alpha\in\Omega_D(M_n(\mathbb{C})).$$

Consider $\nabla^2:\mathscr{H}\to\mathscr{H}\otimes_{M_n(\mathbb{C})}\Omega^2_D(M_n(\mathbb{C}))$. One has $\nabla^2(\phi B) = (\nabla^2\phi)B$, $\forall\phi\in\mathscr{H}$, $\forall B\in M_n(\mathbb{C})$, i.e., ∇^2 is a right-module homomorphism which is *the curvature of* ∇. One defines the component φ of ∇^2 in gauge e by $\nabla^2 e = e\otimes\varphi$; $\varphi\in\Omega^2_D(M_n(\mathbb{C}))$ is given by $\varphi = d\alpha + (\alpha)^2$ where α is the component of ∇ in e. Under a gauge transformation $e\mapsto eU$, $(U\in U(n))$, φ transforms homogeneously as $\varphi\mapsto U^{-1}\varphi U$.

Pure gauge connections have vanishing curvature; indeed one has $\overset{(e)}{\nabla} e = 0$, $\overset{(e)}{\nabla}(eB) = e\otimes dB$ so $\overset{(e)}{\nabla}{}^2(eB) = (\overset{(e)}{\nabla} e)dB + e\otimes d^2B = 0$, $\forall B\in M_n(\mathbb{C})$.

Formula (6′) reads $d(-i\theta) + (-i\theta)^2 = 0$, which is equivalent to $\overset{0}{\nabla}{}^2 = 0$. Thus $\overset{0}{\nabla}$ has a vanishing curvature.

If ∇ is a gauge invariant connection, $\nabla\phi = -i\phi\otimes(\theta + \lambda_k\theta^k)$ for $\lambda_k\in\mathbb{C}$ in view of the Lemma. Then $\nabla^2\phi = -i\phi\otimes d(\lambda_k\theta^k) = i\phi\otimes C^k_{lm}\lambda_k\theta^l\theta^m$ so $\nabla^2=0$ is equivalent to $\lambda_k=0$, i.e., to $\nabla = \overset{0}{\nabla}$. Thus $\overset{0}{\nabla}$ is the only gauge invariant connection with vanishing curvature. One has, in fact, the following stronger result.

Proposition: Let ∇ be a Hermitian connection on $\mathscr{H}$ with vanishing curvature (i.e., $\nabla^2=0$). Then either ∇ is a pure gauge connection $\overset{(e)}{\nabla}$ or ∇ is the canonical connection $\overset{0}{\nabla}$.

Proof: It is sufficient to work in a gauge e. Then $\nabla e = e\otimes\alpha$ with $\alpha = -\bar\alpha$ and $\nabla^2 e = e\otimes\varphi$ is equivalent to $d\alpha + (\alpha)^2 = \varphi$. Setting $\alpha = \beta - i\theta$, the last equation reads $d\beta + \beta^2 - i(\theta\beta + \beta\theta) = \varphi$, i.e.,

$$i[E_k,B_l] - i[E_l,B_k] - B_m C^m_{kl} + [B_k,B_l] - i[E_k,B_l] - i[B_k,E_l] = F_{kl},$$

where $\beta = B_l\theta^l$ and $\varphi = \frac{1}{2}F_{kl}\theta^k\theta^l$. Thus one has

$$\varphi = \tfrac{1}{2}([B_k,B_l] - C^m_{lk}B_m)\theta^k\theta^l. \qquad (8)$$

Therefore, $\varphi=0$ is equivalent to $[B_k,B_l] = C^m_{kl}B_m$ so the $n\times n$ anti-Hermitian matrices B_k satisfies the commutation relations of a basis of SU(n), thus either the corresponding representation of SU(n) is trivial, i.e., $B_k = 0\,\forall k\in\{1,\ldots,n^2-1\}$ which means $\nabla = \overset{0}{\nabla}$ or the representation is unitarily equivalent to the fundamental representation of SU(n), i.e., $B_k = U^{-1}(iE_k)U$, $\forall k\in\{1,\ldots,n^2-1\}$ for some $U\in U(n)$, that reads $\alpha = U^{-1}(i\theta)U - i\theta = U^{-1}dU$, which means $\nabla = \overset{(eU^{-1})}{\nabla}$. □

D. The analog of Maxwell action

Let ∇ be a Hermitian connection on $\mathscr{H}$, let α be its component in a gauge e, and let $\varphi = d\alpha + \alpha^2\in\Omega^2_D(M_n(\mathbb{C}))$ be the component of its curvature ∇^2 in e. The expression $\langle\varphi|\varphi\rangle = -\int\varphi.(*\varphi)$, ($\bar\varphi = -\varphi$ from hermiticity of ∇), is independent of e. We denote it by $\|\nabla^2\|^2 = \langle\varphi|\varphi\rangle$. This notation is clearly justified and $\frac{1}{2}\|\nabla^2\|^2$ is the analog of the classical action of the electromagnetic field. Writing $\alpha = \beta - i\theta$ one has a view of (8):

$$\|\nabla^2\|^2 = -\frac{1}{8n}\sum_{k,l}\mathrm{Tr}\{([B_k,B_l] - C^m_{kl}B_m)^2\},$$

$\|\nabla^2\|^2\geqslant 0$ and its minima correspond to $\nabla^2=0$ and consist of two distinct gauge orbits: The pure gauge connections and the gauge invariant connection $\overset{0}{\nabla}$, which is a singular gauge orbit reduced to a point.

VIII. MISCELLANEOUS RESULTS FOR $M_2(\mathbb{C})$

A. Pauli matrices

As basis of Hermitian traceless 2×2 matrices, we take the Pauli matrices:

$$\sigma_1 = \begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \sigma_2 = \begin{pmatrix}0&-i\\i&0\end{pmatrix},\quad \sigma_3 = \begin{pmatrix}1&0\\0&-1\end{pmatrix}.$$

One has: $\sigma_k\sigma_l = \delta_{kl}\mathbf{1} + i\sum_m\epsilon_{klm}\sigma_m$. Thus comparing to (1), one obtains $g_{kl} = \delta_{kl}$, $C^m_{kl} = C_{klm} = -2\epsilon_{klm}$ and $s^m_{kl} = 0$.

As in Sec. III, one introduces $\partial_k = \mathrm{ad}(i\sigma_k)$ and the $\theta^k\in\Omega^1_D(M_2(\mathbb{C}))$ satisfying $\theta^k(\partial_l) = \delta^k_l\mathbf{1}$. In this case, it is easy to diagonalize the Laplacian on $\Omega_D(M_2(\mathbb{C}))$.

B. Diagonalization of the Laplacian

Since the Laplacian is invariant in the sense that one has $L_X\Delta = \Delta L_X$, $\forall X\in\mathrm{Der}(M_2(\mathbb{C}))$, it follows that the irreductible components of $\Omega_D(M_2(\mathbb{C}))$ for the representation of sl$(2,\mathbb{C})$ given by $X\mapsto L_X$ are the eigenspaces for Δ. [This representation is a representation of the adjoint representation, i.e., of so(3)]. Furthermore, one has $*\Delta = \Delta*$ so it is sufficient to look at $\Omega^0_D(M_2(\mathbb{C}))$ and $\Omega^1_D(M_2(\mathbb{C}))$.

Here, $\Omega^0_D(M_n(\mathbb{C})) = M_n(\mathbb{C})$ splits into two irreductible components: The one-dimensional subspace spanned by $\mathbf{1}$ and the three-dimensional subspace spanned by the σ_k's. One has $\Delta\mathbf{1}=0$ and $\Delta\sigma_k = 8\sigma_k$.

Then, $\Omega^1_D(M_n(\mathbb{C}))$ splits into four irreductible components: The three-dimensional subspace spanned by the θ^k's, the three-dimensional subspace spanned by the $\sigma_k\theta^l - \sigma_l\theta^k$ or equivalently by the $d\sigma_k$'s, the five-dimensional subspace spanned by the $\sigma_k\theta^l + \sigma_l\theta^k - \frac{2}{3}\delta^k_l\sigma_n\theta^n$ and the one-dimensional subspace spanned by $\theta = \sigma_n\theta^n$. One has $\Delta\theta^k = 4\theta^k$,

$\Delta d\sigma_k = 8d\sigma_k$,

$$\Delta(\sigma_k\theta^l + \sigma_l\theta^k - \tfrac{2}{3}\delta^l_k\sigma_n\theta^n) = 16(\sigma_k\theta^l + \sigma_l\theta^k - \tfrac{2}{3}\delta^l_k\sigma_n\theta^n), \text{ and } \Delta\theta = 4\theta.$$

For $\Omega^2_D(M_n(\mathbb{C}))$ one takes the star of the decomposition of $\Omega^1_D(M_n(\mathbb{C}))$, and for $\Omega^3_D(M_2(\mathbb{C}))$ one takes the star of the decomposition of $\Omega^0_D(M_2(\mathbb{C}))$.

Similar consideration of invariance and commutation with the star isomorphism apply for $M_n(\mathbb{C})$, however, to compute the eigenvalues of Δ one needs explicitly the coefficients in formula (1).

IX. CONCLUSION

We have shown that, in spite of the fact that the derivations of $M_n(\mathbb{C})$ are inner, one may develop a relatively rich differential geometric structure by using $\Omega_D(M_n(\mathbb{C}))$ as algebra of differential forms. It is worth noticing here that in quantum mechanics the derivations are also, in some sense, inner derivations and that the discussion of Sec. IV on symplectic structure is clearly relevant there.

This paper is self-contained. Nevertheless, it is a preliminary for our next paper[9] which will deal with the noncommutative differential geometry of algebras of matrix-valued functions on space-time and new types of gauge theory. The content of Sec. VII will be especially relevant for these developments.

[1] W. Greub. S. Halperin, and R. Vanstone, *Connections, Curvature and Co-*

homology, III (Academic, New York, 1976).
[2]M. Dubois-Violette "Derivations et calcul differentiel non-commutatif," C. R. Acad. Sci. Paris **307** (I), 403 (1988).
[3]A. Connes, "Non-commutative differential geometry," Publi. I.H.E.S. **62**, 257 (1986).
[4]M. Karoubi, "Homologie cyclique des groupes et algèbres," C. R. Acad. Sci. Paris **297** (I), 381 (1983).
[5]C. Chevaley and S. Eilenberg, "Cohomology theory of Lie groups and Lie algebras," Trans. Am. Math. Soc. **63**, 85 (1948).
[6]J. L. Koszul, *Homologie et Cohomologie des Algèbres de Lie*, Ph. D. thesis (Gauthier-Villars, Paris, 1950).
[7]H. Cartan, in "*Colloque de Topologie,*" Bruxelles, 1950 (Masson, Paris, 1951).
[8]A. Connes, "C*-algèbres et géométrie différentielle," C. R. Acad. Sci. Paris **290** (A), 599 (1980).
[9]M. Dubois-Violette, R. Kerner, and J. Madore, "Noncommutative differential geometry and new models of gauge theory," J. Math. Phys. **31**, 323 (1990).

Noncommutative differential geometry and new models of gauge theory

Michel Dubois-Violette
Laboratoire de Physique Théorique et Hautes Energies,[a] *Université de Paris-Sud, Bâtiment 211, F-91405 Orsay Cedex, France*

Richard Kerner
Laboratoire de Physique Théorique des Particules Elementaires, Université Pierre et Marie Curie, 4, place Jussieu, F-75252 Paris Cedex 05, France

John Madore
Laboratoire de Physique Théorique et Hautes Energies,[a] *Université de Paris-Sud, Bâtiment 211, F-91405 Orsay Cedex, France*

(Received 13 April 1989; accepted for publication 26 July 1989)

The noncommutative differential geometry of the algebra $C^\infty(V)\otimes M_n(\mathbb{C})$ of smooth $M_n(\mathbb{C})$-valued functions on a manifold V is investigated. For $n\geqslant 2$, the analog of Maxwell's theory is constructed and interpreted as a field theory on V. It describes a U(n)–Yang–Mills field minimally coupled to a set of fields with values in the adjoint representation that interact among themselves through a quartic polynomial potential. The Euclidean action, which is positive, vanishes on exactly two distinct gauge orbits, which are interpreted as two vacua of the theory. In one of the corresponding vacuum sectors, the SU(n) part of the Yang–Mills field is massive. For the case $n=2$, analogies with the standard model of electroweak theory are pointed out. Finally, a brief description is provided of what happens if one starts from the analog of a general Yang–Mills theory instead of Maxwell's theory, which is a particular case.

I. INTRODUCTION AND NOTATION

Let V be a smooth manifold and let $C^\infty(V)$ be the algebra of smooth complex functions on V considered as an abstract commutative *-algebra. Given a smooth complex vector bundle E on V, one denotes by $\Gamma(E)$ the space of smooth sections of E. This $\Gamma(E)$ is a finite projective $C^\infty(V)$ module. The correspondence $E\to\Gamma(E)$ is an equivalence of the category of smooth complex vector bundles on V with the category of finite projective $C^\infty(V)$ modules. There is a notion of connection on finite projective $C^\infty(V)$ modules that corresponds to the notion of connection on vector bundles. To define it, it is convenient to use the graded differential algebra $\Omega(V)$ of complex differential forms on V. The Lie algebra of complex vector fields on V can be identified with the Lie algebra $\mathrm{Der}(C^\infty(V))$ of derivations of $C^\infty(V)$.

In noncommutative differential geometry, the role of $C^\infty(V)$ is played by a noncommutative associative algebra $\mathscr{A}$.[1,2] Modules of sections of vector bundles are replaced by finite projective $\mathscr{A}$ modules.[1,2] In order to define connections on $\mathscr{A}$ modules and more generally to define noncommutative generalization of differential calculus, one needs a generalization of differential forms. There are several noncommutative generalizations of the de Rham complex.[2–4] Here, as in Ref. 5, we use as a generalization of the algebra of differential forms for $\mathscr{A}$ the graded differential algebra $\Omega_D(\mathscr{A})$ introduced in Ref. 4. We now recall the construction of $\Omega_D(\mathscr{A})$.

Let $\mathrm{Der}(\mathscr{A})$ be the Lie algebra of derivations of $\mathscr{A}$. This is a generalization of the Lie algebra of vector fields. Recall that a p cochain ω on the Lie algebra $\mathrm{Der}(\mathscr{A})$ with values in $\mathscr{A}$ is a p-linear antisymmetric mapping of $\mathrm{Der}(\mathscr{A})$ in $\mathscr{A}$, i.e., a linear mapping $\omega\colon \wedge^p \mathrm{Der}(\mathscr{A})\to\mathscr{A}$. The space of p cochains of $\mathrm{Der}(\mathscr{A})$ with values in $\mathscr{A}$ is denoted by $C^p(\mathrm{Der}(\mathscr{A}),\mathscr{A})$. The direct sum

$$C(\mathrm{Der}(\mathscr{A}),\mathscr{A}) = \bigoplus_{p\in\mathbb{N}} C^p(\mathrm{Der}(\mathscr{A}),\mathscr{A})$$

is naturally a graded algebra. It is a graded differential algebra with differential d defined by

$$d\omega(X_0,X_1,\dots,X_p) = \sum_{0\leqslant k\leqslant p} (-1)^k X_k\,\omega(X_0,\overset{k}{\check{\dots}},X_p) + \sum_{0\leqslant r<s\leqslant p} (-1)^{r+s}\omega([X_r,X_s],X_0,\overset{rs}{\check{\dots}},X_p),$$

for $\omega\in C^p(\mathrm{Der}(\mathscr{A}),\mathscr{A})$ and $X_0,X_1,\dots,X_p\in\mathrm{Der}(\mathscr{A})$. One has

$$\mathscr{A} = C^0(\mathrm{Der}(\mathscr{A}),\mathscr{A})\subset C(\mathrm{Der}(\mathscr{A}),\mathscr{A})$$

and the graded differential algebra $\Omega_D(\mathscr{A})$ is defined to be the smallest differential subalgebra of $C(\mathrm{Der}(\mathscr{A}),\mathscr{A})$ that contains $\mathscr{A}$. Any element of $\Omega^p_D(\mathscr{A})$ is a sum of elements of the form $A_0\,dA_1\cdots dA_p$ with $A_0,A_1,\dots,A_p\in\mathscr{A}$. Here $\Omega_D(C^\infty(V))$ coincides with the graded differential algebra $\Omega(V)$ of differential forms on V.

In this paper, we investigate the noncommutative differential geometry of the algebra $\mathscr{A}=C^\infty(V)\otimes M_n(\mathbb{C})$ of smooth $M_n(\mathbb{C})$-valued functions on a connected, simply connected manifold V. Some aspects of the noncommutative geometry of algebras of that type were investigated in Ref. 6 in a different context. We use $\Omega_D(\mathscr{A})$ as the analog of the differential algebra of exterior forms. We show in Sec. II that

$$\Omega_D(\mathscr{A}) = \Omega_D(C^\infty(V))\otimes\Omega_D(M_n(\mathbb{C})).$$

The second factor $\Omega_D(M_n(\mathbb{C}))$ was investigated in Ref. 5. We introduce in Sec. III the analog of a metric for $\mathscr{A}$ and the corresponding scalar product on $\Omega_D(\mathscr{A})$. In Sec. IV, we

[a] Laboratory associated with Centre National de la Recherche Scientifique.

 0022-2488/90/020323-08$03.00

study connections on the free Hermitian $\mathscr{A}$ modules. It is shown that, for $n \geqslant 2$, there are several gauge orbits of flat connections.

In Sec. V, V is the $(s+1)$-dimensional Euclidean space-time $\mathbb{R}^{s+1}$ and we describe the analog for $\mathscr{A}$ of the Maxwell action. This is an action for connections on the free Hermitian $\mathscr{A}$ module of rank 1. We interpret the corresponding theory in terms of a field theory on space-time. It consists of a $U(n)$–Yang–Mills field minimally coupled to a set of scalar fields with value in the adjoint representation that interact among themselves through a quartic polynomial potential. The Euclidean action, which is positive, vanishes on two distinct gauge orbits. These are interpreted as two vacua for the corresponding quantum field theory. In one of the corresponding vacuum sectors, the $SU(n)$ part of the Yang–Mills field is massive. This sector is the most natural one from the point of view of the noncommutative geometry of $\mathscr{A}$ since the vacuum there corresponds to the pure gauge connections, i.e., the pure gauge noncommutative Maxwell potentials. For the case $n=2$ we discuss the analogies and the differences with the standard model of the electroweak interactions (see, for example, Ref. 7). Finally, we describe the analog for $\mathscr{A}$ of the $U(r)$–Yang–Mills action. It is an action for connections on the free Hermitian $\mathscr{A}$ module of range r. In Sec. VI we present our conclusions.

II. DIFFERENTIAL CALCULUS FOR $\mathscr{A}=C^\infty(V)\otimes M_n(\mathbb{C})$

A. The Lie algebra Der($\mathscr{A}$)

We have that $\mathscr{A}=C^\infty(V)\otimes M_n(\mathbb{C})$ and $M_n(\mathbb{C})$ are naturally *-algebras with units. Associated with any point $x\in V$, there is a homomorphism $\gamma_x: \mathscr{A}\to M_n(\mathbb{C})$ of *-algebras with units defined by $\gamma_x(f\otimes M)=f(x)M$, $\forall f\in C^\infty(V)$ and $\forall M\in M_n(\mathbb{C})$. This γ_x is the evaluation at $x\in V$. The subalgebra $C^\infty(V)\otimes \mathbb{1}$ of $\mathscr{A}$ is the center of $\mathscr{A}$. The Lie algebra $\mathrm{Der}(\mathscr{A})$ of all derivations of $\mathscr{A}$ is a module over the center $C^\infty(V)\otimes\mathbb{1}$ of $\mathscr{A}$, so it is a $C^\infty(V)$ module. Here $\mathrm{Der}(C^\infty(V))$ is the Lie algebra of smooth vector fields on V and $\mathrm{Der}(M_n(\mathbb{C}))$ is the Lie algebra $\mathrm{sl}(n,c)$.[4,5] It is clear that $(\mathrm{Der}(C^\infty(V))\otimes\mathbb{1})\oplus(C^\infty(V)\otimes\mathrm{Der}(M_n(\mathbb{C})))$ is a Lie subalgebra and a $C^\infty(V)$ submodule of $\mathrm{Der}(\mathscr{A})$. It is, in fact, $\mathrm{Der}(\mathscr{A})$.

B. Lemma 2.1

Lemma 2.1: One has

$$\mathrm{Der}(\mathscr{A})=(\mathrm{Der}(C^\infty(V))\otimes\mathbb{1})\oplus(C^\infty(V)\otimes\mathrm{Der}(M_n(\mathbb{C}))).$$

Proof: Let X be a derivation of $\mathscr{A}$. Then $f\mapsto X(f\otimes\mathbb{1})$ is a $M_n(\mathbb{C})$-valued vector field on V. One has

$$X(f\otimes M)=X((f\otimes\mathbb{1})(\mathbb{1}\otimes M))=X((\mathbb{1}\otimes M)(f\otimes\mathbb{1})),$$

i.e.,

$$X(f\otimes\mathbb{1})\mathbb{1}\otimes M+f\otimes\mathbb{1}\,X(\mathbb{1}\otimes M)$$
$$=\mathbb{1}\otimes M\,X(f\otimes\mathbb{1})+X(\mathbb{1}\otimes M)f\otimes\mathbb{1}$$

and therefore

$$X(f\otimes\mathbb{1})\mathbb{1}\otimes M=\mathbb{1}\otimes M\,X(f\otimes\mathbb{1}),\quad \forall f\in C^\infty(V),\quad \forall M\in M_n(\mathbb{C}).$$

It follows that $X(f\otimes\mathbb{1})$ is in $C^\infty(V)\otimes\mathbb{1}$, $\forall f\in C^\infty(V)$. This shows that the restriction $X\restriction C^\infty(V)\otimes\mathbb{1}$ is in $\mathrm{Der}(C^\infty(V))\otimes\mathbb{1}$. The mapping $M\mapsto\gamma_x(X(\mathbb{1}\otimes M))$ is a derivation of $M_n(\mathbb{C})$, $\forall x\in V$. This implies that the restriction $X\restriction\mathbb{1}\otimes M_n(\mathbb{C})$ is in $C^\infty(V)\otimes\mathrm{Der}(M_n(\mathbb{C}))$. □

C. The graded differential algebra $\Omega_D(\mathscr{A})$

We recall that if Ω_0 and Ω_1 are graded differential algebras with differentials d_0 and d_1, then $\Omega_0\otimes\Omega_1$ is naturally a graded differential algebra if one defines the product by

$$(x\otimes y)(z\otimes t)=(-1)^{rs}xz\otimes yt,\qquad \text{for } x\in\Omega_0,\quad y\in\Omega_1^r,\quad z\in\Omega_0^s,\quad t\in\Omega_1,$$

and the differential d by

$$d(x\otimes y)=d_0x\otimes y+(-1)^p x\otimes d_1y,\quad \forall x\in\Omega_0^p,\quad \forall y\in\Omega_1.$$

It follows from Lemma 2.1 that

$$C(\mathrm{Der}(C^\infty(V)),C^\infty(V))\otimes C(\mathrm{Der}(M_n(\mathbb{C})),M_n(\mathbb{C}))$$

is a graded differential subalgebra of $C(\mathrm{Der}(\mathscr{A}),\mathscr{A})$. On the other hand, $\Omega_D(C^\infty(V))$ is the smallest differential subalgebra of $C(\mathrm{Der}(C^\infty(V)),C^\infty(V))$ that contains $C^\infty(V)$ and $\Omega_D(M_n(\mathbb{C}))$ is the smallest differential subalgebra of $C(\mathrm{Der}\,M_n(\mathbb{C}),M_n(\mathbb{C}))$ that contains $M_n(\mathbb{C})$. Therefore the smallest differential subalgebra $\Omega_D(\mathscr{A})$ of $C(\mathrm{Der}(\mathscr{A}),\mathscr{A})$ that contains $\mathscr{A}=C^\infty(V)\otimes M_n(\mathbb{C})$ is $\Omega_D((C^\infty(V))\otimes\Omega_D(M_n(\mathbb{C}))$. Thus one has

$$\Omega_D(\mathscr{A})=\Omega_D(C^\infty(V))\otimes\Omega_D(M_n(\mathbb{C})).$$

In fact,[4] $\Omega_D(C^\infty(V))$ is the graded differential algebra $\Omega(V)$ of exterior differential forms on V and $\Omega_D(M_n(\mathbb{C}))$ coincides with[4,5]

$$C(\mathrm{Der}(M_n(\mathbb{C})),M_n(\mathbb{C}))=M_n(\mathbb{C})\otimes\wedge\,\mathrm{sl}(n,\mathbb{C})^*.$$

So one has

$$\Omega_D(\mathscr{A})=\Omega(V)\otimes M_n(\mathbb{C})\otimes\wedge\,\mathrm{sl}(n,\mathbb{C})^*.$$

D. Remark

Remark 2.2: For algebras $\mathscr{B}$ and $\mathscr{C}$, $\Omega_D(\mathscr{B}\otimes\mathscr{C})$ is generally distinct from $\Omega_D(\mathscr{B})\otimes\Omega_D(\mathscr{C})$. For instance,

$$\Omega_D(M_r(\mathbb{C}))=M_r(\mathbb{C})\otimes\wedge\,\mathrm{sl}(r,\mathbb{C})^*,$$
$$\Omega_D(M_s(\mathbb{C}))=M_s(\mathbb{C})\otimes\wedge\,\mathrm{sl}(s,\mathbb{C})^*,$$
$$\Omega_D(M_r(\mathbb{C})\otimes M_s(\mathbb{C}))=M_r(\mathbb{C})\otimes M_s(\mathbb{C})\otimes\wedge\,\mathrm{sl}(rs,\mathbb{C})^*.$$

In fact, one has

$$\mathrm{sl}(rs,\mathbb{C})=(\mathrm{sl}(r,\mathbb{C})\otimes\mathrm{sl}(s,\mathbb{C}))\oplus(\mathrm{sl}(r,\mathbb{C})\oplus\mathbb{1})\otimes(\mathbb{1}\otimes\mathrm{sl}(s,\mathbb{C})).$$

E. $\Omega_D(\mathscr{A})$ as bigraded differential algebra

Now $\Omega_D(\mathscr{A})$ is naturally a bigraded algebra if one sets

$$\Omega_D^{r,s}(\mathscr{A})=\Omega^r(V)\otimes\Omega_D^s(M_n(\mathbb{C})).$$

We identify $\Omega(V)$ [resp. $\Omega_D(M_n(\mathbb{C}))$] with the differential subalgebra $\Omega(V)\otimes\mathbb{1}$ [resp. $\mathbb{1}\otimes\Omega_D(M_n(\mathbb{C}))$] of $\Omega_D(\mathscr{A})$. We denote by d the differential of $\Omega_D(\mathscr{A})$. Let d' be the unique antiderivation of $\Omega_D(\mathscr{A})$ extending the exterior differential of $\Omega(V)$ such that $d'\Omega_D(M_n(\mathbb{C}))=0$ and let d'' be the unique antiderivation of $\Omega_D(\mathscr{A})$ extending the differential of $\Omega_D(M_n(\mathbb{C}))$ such that $d''\Omega(V)=0$. Then d' is of bidegree (1,0), d'' is of bidegree (0,1), and one has $d=d'+d''$, $d'^2=d''^2=d''d'+d'd''=0$. In other words, $\Omega_D(\mathscr{A})$ is a bigraded differential algebra.

F. Reality

Now $\mathscr{A}$ is an *-algebra. We denote by $\mathrm{Der}_{\mathbb{R}}(\mathscr{A})$ the real Lie subalgebra of $\mathrm{Der}(\mathscr{A})$ of derivations X such that $X(A^*)=(XA)^*$. One has

$$\mathrm{Der}_{\mathbb{R}}(\mathscr{A})=(\mathrm{Der}_{\mathbb{R}}(C^\infty(V))\otimes 1)\oplus(C^\infty_{\mathbb{R}}(V)\otimes\mathrm{Der}_{\mathbb{R}}(M_n(\mathbb{C}))),$$

where $\mathrm{Der}_{\mathbb{R}}(C^\infty(V))$ is the real Lie algebra of real vector fields on V, $\mathrm{Der}_{\mathbb{R}}(M_n(\mathbb{C}))$ is the Lie algebra $\mathbf{su}(n)$ for its adjoint action on $M_n(\mathbb{C})$, and $C^\infty_{\mathbb{R}}(V)$ is the real algebra of real functions on V. Correspondingly, there is an antilinear involution $\omega\mapsto\bar\omega$ on $\Omega_D(\mathscr{A})$ that extends the involution of $\mathscr{A}$. One has $\overline{\alpha\otimes\alpha'}=\bar\alpha\otimes\bar\alpha'$, for $\alpha\in\Omega(V)$ and $\alpha'\in\Omega_D(M_n(\mathbb{C}))$, where $\alpha\mapsto\bar\alpha$ is the complex conjugation of differential forms on V and $\alpha'\mapsto\bar\alpha'$ is the involution of $\Omega_D(M_n(\mathbb{C}))$ defined in Ref. 5. An element ω of $\Omega_D(\mathscr{A})$ will be said to be *real* (resp. *imaginary*) if $\bar\omega=\omega$ (resp. $\bar\omega=-\omega$).

III. METRIC FOR $\mathscr{A}$ AND SCALAR PRODUCT ON $\Omega_D(\mathscr{A})$

A. Basis for $M_n(\mathbb{C})$ and expressions for d''

For the differential calculus of $M_n(\mathbb{C})$ [or $\Omega_D(M_n(\mathbb{C}))$] we use the notation of Ref. 5, except that, since we consider $\Omega_D(M_n(\mathbb{C}))$ to be embedded in $\Omega_D(\mathscr{A})$, the differential of $\Omega_D(M_n(\mathbb{C}))$ will be denoted d'' (or d) to be consistent with Sec. II E. We shall use a basis E_k, $k\in\{1,\dots,n^2-1\}$, of Hermitian traceless $n\times n$ matrices, which is orthonormal in the sense that $(1/n)\mathrm{Tr}(E_kE_l)=\delta_{kl}$. So one has a multiplication table in $M_n(\mathbb{C})$ of the form

$$E_kE_l=\delta_{kl}\mathbf{1}+\sum_m\left(s_{klm}-\frac{i}{2}C_{klm}\right)E_m, \tag{1}$$

with $s_{klm}=s_{lkm}\in\mathbb{R}$ and $C_{klm}=C_{lkm}\in\mathbb{R}$. Associativity then implies that s_{klm} is completely symmetric, that C_{klm} is completely antisymmetric, and that they satisfy some relations (see Ref. 8, for instance). It follows from these relations that one has

$$\sum_{k,l}C_{klr}C_{kls}=2n^2\delta_{rs}.$$

Let us introduce the basis $\partial_k=\mathrm{ad}(iE_k)$ of $\mathrm{Der}_{\mathbb{R}}(M_n(\mathbb{C}))=\mathbf{su}(n)$. One has

$$[\partial_k,\partial_l]=\sum_m C_{klm}\,\partial_m.$$

Define

$$\theta^k\in\Omega^1_D(M_n(\mathbb{C}))\subset\Omega_D(\mathscr{A}),\quad k\in\{1,\dots,n^2-1\},$$

by

$$\theta^k(\partial_l)=\delta^k{}_l\mathbf{1},\quad\text{for}\quad k,l\in\{1,\dots,n^2-1\}.$$

One has, in $\Omega_D(\mathscr{A})$,

$$A\theta^k=\theta^kA,\quad\forall A\in\mathscr{A}, \tag{2}$$

and

$$\theta^k\theta^l=-\theta^l\theta^k. \tag{3}$$

The differential d'' is then characterized by

$$d''\alpha=0,\quad\forall\alpha\in\Omega(V), \tag{4}$$

$$d''E_k=-\sum_{m,l}C_{klm}E_m\theta^l, \tag{5}$$

and

$$d''\theta^k=-\frac{1}{2}\sum_{l,m}C_{klm}\theta^l\theta^m. \tag{6}$$

By introducing the canonical element θ of $\Omega^1_D(M_n(\mathbb{C}))$,[5] defined by $\theta=E_k\theta^k$ and using Eq. (4), Eq. (5) can be rewritten in the form

$$d''A=i[\theta,A],\quad\forall A\in\mathscr{A}. \tag{5'}$$

Relation (6) may be inverted to yield

$$\theta^k=-\frac{i}{n^2}\sum_l E_lE_k\,d''E_l.$$

The differential d' is characterized by the fact that it coincides with the exterior differential on $\Omega(V)\subset\Omega_D(\mathscr{A})$ and that it satisfies $d'E_k=0$ and $d'\theta^k=0$, for $k\in\{1,\dots,n^2-1\}$.

B. Metric for $\mathscr{A}$

We now asume that V is an oriented Riemannian manifold with metric ds^2. In local coordinates (x^μ), $ds^2=g_{\mu\nu}\,dx^\mu\otimes dx^\nu$, and $(g^{\mu\nu})$ will denote the inverse matrix of $(g_{\mu\nu})$.

In Ref. 5, we introduced what we called there the canonical Riemannian structure for $M_n(\mathbb{C})$, which becomes, with conventions adopted here,

$$\sum_k\theta^k\otimes\theta^k.$$

It is natural to combine these structures by introducing the metric

$$ds^2+\left(\frac{1}{\mathbf{m}}\right)^2\sum_k\theta^k\otimes\theta^k,$$

for $\mathscr{A}$, where $1/\mathbf{m}$ is a positive constant.

We have in mind the case where $V=\mathbf{R}^{s+1}$ is the $(s+1)$-dimensional Euclidean space-time and where

$$ds^2=\sum_\mu dx^\mu\otimes dx^\mu$$

has the dimension of the square of a length. In this case $1/\mathbf{m}$ is a length, i.e., $\mathbf{m}$ is a mass in standard units where $\hbar=c=1$.

C. Scalar product for $\Omega_n(\mathscr{A})$

Associated with the metric and the orientation of V, there is the star isomorphism $\alpha\mapsto{}^*\alpha$ of $\Omega(V)$ and the corresponding positive Hermitian scalar product on $\Omega(V)$ such that

$$\langle\alpha|\beta\rangle = \int_V \bar{\alpha}\Lambda^*\beta, \quad \text{for } \alpha,\beta\in\Omega^p(V),$$

and $\langle\alpha|\beta\rangle = 0$, if α and β have different degrees. Strictly speaking, this scalar product is defined on $\Omega(V)$ only if V is compact. Otherwise one has to restrict attention to forms for which $\langle\alpha|\alpha\rangle < \infty$, for instance, to forms with compact support. However, we shall not be concerned with this here since the scalar product will be used only to write formal actions for Euclidean field theories.

In Ref. 5, we constructed a star isomorphism of $\Omega_D(M_n(\mathbb{C}))$ associated to $\Sigma_k\theta^k\otimes\theta^k$ and then defined a scalar product on $\Omega_D(M_n(\mathbb{C}))$ by using this star isomorphism and a generalization of integration (essentially the trace). The only thing that the rescaling

$$\sum_k \theta^k\otimes\theta^k \mapsto \left(\frac{1}{\mathbf{m}}\right)^2 \sum_k \theta^k\otimes\theta^k$$

changes is the scalar product $\langle\alpha''|\beta''\rangle$ of $\alpha'',\beta''\in\Omega_D^p(M_D(\mathbb{C}))$. It becomes $(1/\mathbf{m})^{n^2-1}\times\mathbf{m}^{2p}$ times the scalar product of Ref. 5, which corresponds to the case $\mathbf{m}=1$.

We now define a scalar product $\langle\cdot|\cdot\rangle$ on $\Omega_D(\mathscr{A})$ by

$$\langle\alpha'\otimes\alpha''|\beta'\otimes\beta''\rangle = \langle\alpha'|\beta'\rangle\langle\alpha''|\beta''\rangle, \quad \forall\alpha',\beta'\in\Omega(V),$$
$$\forall\alpha'',\beta''\in\Omega_D(M_n(\mathbb{C})).$$

This is just the scalar product we would obtain from

$$ds^2 + \left(\frac{1}{\mathbf{m}}\right)^2 \sum_k \theta^k\otimes\theta^k$$

by proceeding as in Ref. 5.

IV. CONNECTIONS ON HERMITIAN $\mathscr{A}$ MODULES

A. Hermitian $\mathscr{A}$ modules

An element P of $\mathscr{A}$ is *positive* if $P = A^*A$, for some $A\in\mathscr{A}$. The set $\mathscr{A}^+$ of positive elements of $\mathscr{A}$ is a convex cone in $\mathscr{A}$. Let $\mathscr{M}$ be a right $\mathscr{A}$ module. A Hermitian structure on $\mathscr{M}$ is a $\mathscr{A}$-valued positive-definite Hermitian form on $\mathscr{M}$,

$$(\Psi,\Phi)\mapsto h(\Psi,\Phi)\in\mathscr{A} \quad (\Psi,\Phi\in\mathscr{M}),$$

such that one has

$$h(\Psi A,\Phi B) = A^* h(\Psi,\Phi)B, \quad \forall\Psi,\Phi\in\mathscr{M}, \quad \forall A,B\in\mathscr{A}.$$

Positive definite means that $h(\Psi,\Psi)\in\mathscr{A}^+$, $\forall\Psi\in\mathscr{M}$, and that $h(\Psi,\Psi)=0$ implies $\Psi=0$.

A right $\mathscr{A}$ module equipped with a Hermitian structure will be called a Hermitian $\mathscr{A}$ module.

Now $\mathscr{A}^r$ is naturally a right $\mathscr{A}$ module:

$$(A_1,\dots,A_r)A = (A_1A,\dots,A_rA),$$
$$\forall(A_1,\dots,A_r)\in\mathscr{A}^r, \quad \forall A\in\mathscr{A}.$$

It is a Hermitian $\mathscr{A}$ module if one defines its Hermitian structure by

$$h((A_1,\dots,A_r),(B_1,\dots,B_r)) = \sum_{a=1}^{a=r} A^*_a B_a.$$

Conversely, let $\mathscr{H}^r$ be a free Hermitian $\mathscr{A}$ module of rank r with Hermitian structure h. Then one can construct an orthonormal basis (e_a), $a\in\{1,\dots,r\}$ of $\mathscr{H}^r$, i.e., $e_a\in\mathscr{H}^r$ such that

$$h(e_a,e_b) = \delta_{ab}\mathbf{1}, \quad \forall a,b\in\{1,\dots,r\}.$$

We shall call such an orthonormal basis *a gauge*. Given such a gauge, $\Psi\in\mathscr{H}^r$ can be written

$$\Psi = \sum_a e_a A_a$$

in a unique way with $A_a\in\mathscr{A}$.

Furthermore if

$$\Phi = \sum e_a B_a$$

is another element of $\mathscr{H}^r$, then

$$h(\Psi,\Phi) = \sum_a A^*_a B_a.$$

Thus each gauge gives an isomorphism $\mathscr{H}^r\to\mathscr{A}^r$ of Hermitian $\mathscr{A}$ modules. A change of orthonormal basis will be called a *gauge transformation*. Such a gauge transformation U is a unitary element of

$$\mathscr{A}\otimes M_r(\mathbb{C}) = C^\infty(V)\otimes M_n(\mathbb{C})\otimes M_r(\mathbb{C})$$
$$= C^\infty(V)\otimes M_{nr}(\mathbb{C}).$$

So U is a $U(nr)$-valued function on V.

B. Connections

Let $\mathscr{M}$ be a right $\mathscr{A}$ module. A Ω_D *connection* or simply a *connection* on $\mathscr{M}$ (see Ref. 2) is a linear mapping. $\nabla: \mathscr{M}\to\mathscr{M}\otimes_{\mathscr{A}}\Omega_D^1(\mathscr{A})$ such that

$$\nabla(\Phi A) = (\nabla\Phi)A + \Phi\otimes dA, \quad \forall\Phi\in\mathscr{M}, \quad \forall A\in\mathscr{A}.$$

If $\mathscr{M}$ is a Hermitian $\mathscr{A}$ module with Hermitian structure h, ∇ will be called a *Hermitian connection* if it satisfies

$$dh(\Phi,\Psi) = h(\nabla\Phi,\Psi) + h(\Phi,\nabla\Psi), \quad \forall\Phi,\Psi\in\mathscr{M}.$$

Connections always exist on projective modules of finite type.[2]

Let ∇ be a connection on $\mathscr{M}$. One extends ∇ as a linear mapping, again denoted by ∇, of $\mathscr{M}\otimes_{\mathscr{A}}\Omega_D(\mathscr{A})$ in itself setting[2]

$$\nabla(\Phi\otimes\alpha) = (\nabla\Phi)\alpha + \Phi\otimes d\alpha, \quad \forall\Phi\in\mathscr{M}, \quad \forall\alpha\in\Omega_D(\mathscr{A}).$$

Consider

$$\nabla^2: \mathscr{M}\to\mathscr{M}\otimes_{\mathscr{A}}\Omega_D^2(\mathscr{A}).$$

One has

$$\nabla^2(\Phi A) = (\nabla^2\Phi)A, \quad \forall\Phi\in\mathscr{M}, \quad \forall A\in\mathscr{A}.$$

Thus ∇^2 is a right $\mathscr{A}$ module homomorphism which is *the curvature of* ∇.

C. Connections on the free Hermitian $\mathscr{A}$ module of rank r

We consider $\mathscr{A}^r$ as a Hermitian right $\mathscr{A}$ module as explained in Sec. IV A. The canonical basis of $\mathscr{A}^r$ will be denoted by $e=(e_1,...,e_r)$. We denote by $\mathscr{U}_r$ the group of gauge transformations, i.e., the group of unitary elements of $M_r(\mathscr{A})=\mathscr{A}\otimes M_r(\mathbb{C})$. Any orthonormal basis, or gauge, of $\mathscr{A}^r$ is of the form eU for a unique $U\in\mathscr{U}_r$.

Let ∇ be a connection on $\mathscr{A}^r$. Then $\nabla e_a = e_b\otimes\omega^b_a$, for $a\in\{1,...,r\}$, $\omega^a_b\in\Omega^1_D(\mathscr{A})$. Furthermore, ∇ is Hermitian if and only if $\overline{\omega}^a_b = -\omega^b_a$. We write the relations $\nabla e_a = e_b\omega^b_a$ in the form $\nabla e = e\omega$ with

$$\omega = (\omega^a_b)\in M_r(\Omega^1_D(\mathscr{A})) = \Omega^1_D(\mathscr{A})\otimes M_r(\mathbb{C}).$$

The element ω of $\Omega^1_D(\mathscr{A})\otimes M_r(\mathbb{C})$ will be called *the component of* ∇ *in* e or simply the component of ∇. Each $\omega\in\Omega^1_D(\mathscr{A})\otimes M_r(\mathbb{C})$ is the component in e of a unique connection ∇. We could define similarly the component of ∇ in an arbitrary gauge eU: If ω is its component in e, then its component in eU is $U^{-1}\omega U + U^{-1}\,dU$. Here, however, we consider $U^{-1}\omega U + U^{-1}\,dU$ as the component in e of another connection denoted ∇^U. The $\nabla\mapsto\nabla^U$, $U\in\mathscr{U}_r$ is a right action of the gauge group $\mathscr{U}_r$ on the space of connections on $\mathscr{A}^r$. The connection ∇^U is Hermitian if and only if ∇ is Hermitian. The set $\{\nabla^U | U\in\mathscr{U}_r\}$ will be called *the gauge orbit of* ∇. In the same way, $\nabla^2 e = e\varphi$, with $\varphi\in\Omega^2_D(\mathscr{A})\otimes M_r(\mathbb{C})$. One has $\varphi = d\omega+\omega^2$ in the algebra $\Omega_D(\mathscr{A})\otimes M_r(\mathbb{C})$, where d is defined by

$$d(\alpha\otimes x) = d\alpha\otimes x,\quad \forall\alpha\in\Omega_D(\mathscr{A}),\quad \forall x\in M_r(\mathbb{C}).$$

Here φ will be called *the component of the curvature* ∇^2 of ∇. If φ is the component of ∇^2, then the component of the curvature $(\nabla^U)^2$ of ∇^U is $U^{-1}\varphi U$.

D. Flat Hermitian connections on $\mathscr{A}^r$

A connection is called a flat connection if its curvature vanishes. Thus a connection ∇ on $\mathscr{A}^r$ with component ω is flat if and only if $d\omega+\omega^2=0$. If $U\in\mathscr{U}_r$, then ∇^U is flat if and only if ∇ is flat.

For each gauge eU^{-1} ($U\in\mathscr{U}_r$) there is a unique connection $\overset{(eU^{-1})}{\nabla}$ such that $\overset{(eU^{-1})}{\nabla}(eU^{-1})=0$. Its component in e is $U^{-1}\,dU$ so one has $\overset{(eU^{-1})}{\nabla} = \overset{(e)}{\nabla}{}^U$ and it is a flat Hermitian connection. These connections $\overset{(e)}{\nabla}{}^U$, $U\in\mathscr{U}_r$, will be called *pure gauge connections*. The set of pure gauge connections is a gauge orbit of flat Hermitian connections on $\mathscr{A}^r$. In the commutative case where $\mathscr{A}=C^\infty(V)$ they are the only flat Hermitian connections on $\mathscr{A}^r$. However, for $\mathscr{A}=C^\infty(V)\otimes M_n(\mathbb{C})$, with $n\geqslant 2$, there are other gauge orbits of Hermitian flat connections on $\mathscr{A}^r$, which we now describe.

We now assume that $\mathscr{A}=C^\infty(V)\otimes M_n(\mathbb{C})$, with $n\geqslant 2$, and we let r denote a positive integer with $r\geqslant 1$. Let R^α_k, $k\in\{1,2,...,n^2-1\}$, $\alpha\in\{0,1,...,N(n,r)\}$, be a set of anti-Hermitian elements of $M_n(\mathbb{C})\otimes M_r(\mathbb{C})$ such that

$$R^0_k = 0,\quad R^1_k = iE_k\otimes 1,\quad [R^\alpha_k, R^\alpha_l] = \sum_m C_{klm}R^\alpha_m$$

[i.e., R^α is a representation of $\mathbf{su}(n)$ in $\mathbb{C}^n\otimes\mathbb{C}^r$], $\forall\alpha,k,l$, and such that, if (R_k) are n^2-1 anti-Hermitian elements of $M_n(\mathbb{C})\otimes M_r(\mathbb{C})$ satisfying

$$[R_k,R_l] = \sum_m C_{klm}R_m,\quad \forall k,l,$$

then there is a unique $\alpha\in\{0,1,...,N(n,r)\}$ and a unitary $V\in M_n(\mathbb{C})\otimes M_r(\mathbb{C})$ such that $R_k = V^{-1}R^\alpha_kV$, $\forall k$.

In other words, (R^α) is a complete set of mutually inequivalent anti-Hermitian representations of $\mathbf{su}(n)$ in $\mathbb{C}^n\otimes\mathbb{C}^r$. Let $\overset{\alpha}{\nabla}$ be the connection on $\mathscr{A}^r$ with component

$$(R^\alpha_k - iE_k\otimes 1)\theta^k\in\Omega^1_D(\mathscr{A})\otimes M_r(\mathbb{C}),\qquad \forall\alpha\in\{0,1,...,N(n,r)\}.$$

The $\overset{\alpha}{\nabla}$ are Hermitian connections and one has the following result.

E. Theorem 4.1

Theorem 4.1: (a) The $\overset{\alpha}{\nabla}$ are flat Hermitian connections and, if α is distinct of β, the gauge orbits of $\overset{\alpha}{\nabla}$ and of $\overset{\beta}{\nabla}$ are distinct.

(b) A Hermitian connection ∇ on $\mathscr{A}^r$ is flat if and only if it is an element of the gauge orbit of $\overset{\alpha}{\nabla}$ for some $\alpha\in\{0,1,...,N(n,r)\}$, i.e., $\nabla=\overset{\alpha}{\nabla}{}^U$ with $U\in\mathscr{U}_r$ and $\alpha\in\{0,...,N(n,r)\}$.

Proof: Let ∇ be a Hermitian connection on $\mathscr{A}^r$ with component ω. Write ω in the form

$$\omega = A + (B_k - iE_k\otimes 1)\theta^k,$$

where A is a one-form on V with values in the anti-Hermitian elements of $M_n(\mathbb{C})\otimes M_r(\mathbb{C})$ and where the B_k are functions on V with values in the anti-Hermitian elements of $M_n(\mathbb{C})\otimes M_r(\mathbb{C})$. One has

$$d\omega+\omega^2 = d'A + A^2 + (d'B_k + [A,B_k])\theta^k + \frac{1}{2}\Big([B_k,B_l] - \sum_m C_{klm}B_m\Big)\theta^k\theta^l. \quad (7)$$

Here ∇ is flat if and only if

$$d'A + A^2 = 0,\quad d'B_k + [A,B_k] = 0,\quad \forall k,$$

and

$$[B_k,B_l] = \sum_m C_{klm}B_m,\quad \forall k,l.$$

It follows that the $\overset{\alpha}{\nabla}$ are flat connections. If $\overset{\beta}{\nabla} = \overset{\alpha}{\nabla}{}^U$, U may be chosen to be constant and then $R^\beta_k = U^{-1}R^\alpha_kU$, which is in contradiction with the assumptions on the R^α.

Suppose that ∇ is a flat Hermitian connection. Then $d'A + A^2 = 0$ implies $A = U^{-1}d'U$ and

$$[B_k,B_l] = \sum_m C_{klm}B_m$$

implies $B_k = V^{-1}R^\alpha_kV$ for some $\alpha\in\{0,1,...,N(n,r)\}$ and $U,V\in\mathscr{U}_r$. Furthermore $d'B_k + [A,B_k] = 0$ implies $d'(UV^{-1}R^\alpha_kVU^{-1}) = 0$, so one can choose U and V such that $U=V$. This implies $\nabla = \overset{\alpha}{\nabla}{}^U$. □

F. Remarks

(a) Under a gauge transformation $\nabla \mapsto \nabla^U$, $U \in \mathscr{U}_r$, A and B_k as above transform as $A \mapsto U^{-1}AU + U^{-1}d'U$ and $B_k \mapsto U^{-1}B_k U$. Thus the B_k transform homogeneously. This is, in fact, the reason why we represent the component ω of ∇ in the form $\omega = A + (B_k - iE_k \otimes \mathbb{1})\theta^k$ and why we introduce the component of $\overset{\alpha}{\nabla}$ in the form $(R_k^\alpha - iE_k \otimes \mathbb{1})\theta^k$. It is connected with what was described in Ref. 5, Lemma 7.3, for matrix algebras.

(b) One has $\overset{1}{\nabla} = \overset{(e)}{\nabla}$ so the pure gauge connections on $\mathscr{A}^r$ are the elements of the gauge orbit of $\overset{1}{\nabla}$.

(c) For any $r \geqslant 1$, $N(n,r) \geqslant 1$ $(n \geqslant 2)$, so one has at least two gauge orbits of flat Hermitian connections of $\mathscr{A}^r$: The orbit of $\overset{0}{\nabla}$ and the orbit of $\overset{1}{\nabla}$, which is the set of pure gauge connections. In the case $r = 1$, $N(n,1) = 1$, so one has only these two gauge orbits.

(d) Formulas like (7) naturally appear in the double-bundle structures (see, for example, Ref. 9).

V. MODELS OF GAUGE THEORY

A. Classical Euclidean Maxwell and Yang–Mills actions

Throughout Sec. V, $V = \mathbb{R}^{s+1}$ is the $(s+1)$-dimensional Euclidean space-time with metric

$$ds^2 = \sum_{\mu=0}^{\mu=s} (dx^\mu)^2$$

and

$$\mathscr{A} = C^\infty(\mathbb{R}^{s+1}) \otimes M_n(\mathbb{C}).$$

We recall here in the case $n = 1$, i.e., $\mathscr{A} = C^\infty(\mathbb{R}^{s+1})$, the definition of the Maxwell action and, in general, that of the $U(r)$–Yang–Mills action.

The Maxwell action is an action for connections on a $U(1)$ principal bundle over $\mathbb{R}^{s+1}$. One can also say that it is an action for Hermitian connections on a Hermitian vector bundle of rank 1 over $\mathbb{R}^{s+1}$. Finally, since $\mathbb{R}^{s+1}$ is contractible, it is an action for Hermitian connections on the free Hermitian $C^\infty(\mathbb{R}^{s+1})$ module of rank 1. Let ∇ be such a connection with component

$$A = A_\mu\, dx^\mu \in \Omega^1(V) = \Omega_D^1(C^\infty(\mathbb{R}^{s+1}))$$

(the Maxwell potential), and component $F = \frac{1}{2} F_{\mu\nu}\, dx^\mu \wedge dx^\nu$ of the curvature ∇^2 (the corresponding electromagnetic field). One has $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. The Maxwell action $S(\nabla)$ for ∇ is

$$S(\nabla) = \|\nabla^2\|^2 = -\frac{1}{4}\int \sum_{\mu,\nu} (\partial_\mu A_\nu - \partial_\nu A_\mu)^2\, d^{s+1}x. \tag{8}$$

This action is gauge invariant, positive, and vanishes only on the gauge orbit of pure gauge connections. Two connections in the same gauge orbit are considered as physically equivalent.

In the same way, the $U(r)$–Yang–Mills action is an action for Hermitian connections on the free Hermitian $C^\infty(\mathbb{R}^{s+1})$ module of rank r. If ∇ is such a connection with component

$$A = A_\mu\, dx^\mu \in \Omega^1(V) \otimes M_r(\mathbb{C}),$$

the Yang–Mills action is given by

$$S(\nabla) = -\frac{1}{4}\int \sum_{\mu,\nu} \frac{1}{r} \mathrm{Tr}((\partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu])^2)\, d^{s+1}x. \tag{8'}$$

This action is again gauge invariant, positive, and vanishes only if ∇ is a pure gauge connection. It coincides with Maxwell action for $r = 1$.

B. Maxwell action for $\mathscr{A} = C^\infty(\mathbb{R}^{s+1}) \otimes M_n(\mathbb{C})$

It is natural to generalize the Maxwell action for arbitrary positive integer n as $\|\nabla^2\|^2$ on Hermitian connection ∇ on the free Hermitian $\mathscr{A}$ module of rank 1. Let $\omega \in \Omega_D^1(\mathscr{A})$ be the component of ∇; then $\|\nabla^2\|^2$ means $\langle d\omega + \omega^2 | d\omega + \omega^2 \rangle$ with the scalar product defined in Sec. III C on $\Omega_D(A)$.

Since, from Sec. III C, we know that there is an overall scale factor $(1/\mathbf{m})^{n^2-1}$ in front of this scalar product we define the generalized Maxwell action as

$$S(\nabla) = (\mathbf{m})^{n^2-1} \langle d\omega + \omega^2 | d\omega + \omega^2 \rangle.$$

Writing ω again as

$$\omega = A_\mu\, dx^\mu + (B_k - iE_k)\, \theta^k$$

with anti-Hermitian $n \times n$-matrix-valued functions A_μ, $\mu \in \{0,1,\dots,s\}$, and B_k, $k \in \{1,\dots,n^2-1\}$, $S(\nabla)$ is given by

$$S(\nabla) = -\frac{1}{4n}\int \sum_{\mu,\nu} \mathrm{Tr}((\partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu])^2) - \frac{\mathbf{m}^2}{2n}\int \sum_{\lambda,k} \mathrm{Tr}((\partial_\lambda B_k + [A_\lambda, B_k])^2) - \frac{\mathbf{m}^4}{4n}\int \sum_{k,l} \mathrm{Tr}\Big(\Big([B_k, B_l] - \sum_m C_{klm} B_m\Big)^2\Big),$$

which can also be written

$$S(\nabla) = -\int_{\mathbb{R}^{s+1}} \frac{1}{4n} \mathrm{Tr}(F_{\mu\nu} F^{\mu\nu}) + \frac{1}{2n} \mathrm{Tr}((\nabla_\lambda \phi_k)(\nabla^\lambda \phi^k)) + \frac{1}{4n} \sum_{k,l} \mathrm{Tr}\Big([\phi_k, \phi_l] - \mathbf{m} \sum_m C_{klm} \phi_m\Big)^2, \tag{9}$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu] = F^{\mu\nu}, \quad \phi_k = \mathbf{m}\, B_k,$$

$$\nabla_\lambda \phi_k = \partial_\lambda \phi_k + [A_\lambda, \phi_k] = \nabla^\lambda \phi^k.$$

Under a gauge transformation $\nabla \mapsto \nabla^U$, $U \in \mathscr{U}_1$, the A_μ transform as $A_\mu \mapsto U^{-1} A_\mu U + U^{-1} \partial_\mu U$, the ϕ_k transform as $\phi_k \mapsto U^{-1} \phi_k U$, and the $\nabla_\lambda \phi_k$ transform as $\nabla_\lambda \phi_k \mapsto U^{-1}(\nabla_\lambda \phi_k) U$. The action (9) is gauge invariant, positive, and, for $n \geqslant 2$, vanishes (in view of Theorem 4.1) on the gauge orbit of ($A_\mu = 0$, $\phi_k = 0$) and on the gauge orbit of ($A_\mu = 0$, $\phi_k = i\mathbf{m}E_k$).

C. Discussion

The action (9) can be interpreted as the Euclidean action of a field theory on $\mathbb{R}^{s+1}$. It is then the Euclidean action for a $U(n)$–Yang–Mills field minimally coupled to n^2-1 scalar fields ϕ_k with values in the adjoint representation of $U(n)$ that interact among themselves through a quartic polynomial potential.

We now assume that $n \geqslant 2$ and $s+1 \geqslant 2$. Then the two gauge orbits where the action vanishes are separated by an infinite barrier; there is no instanton interpolating between these two gauge orbits. This follows from the translation invariance. Therefore, by standard arguments,[7] each of these orbits corresponds to a vacuum for the corresponding quantum field theory in Minkowski space. Let Ω_0 be the vacuum corresponding to the gauge orbit of ($A_\mu = 0$, $\phi_k = 0$) and let Ω_1 be the one corresponding to the gauge orbit of ($A_\mu = 0$, $\phi_k = i\mathbf{m}E_k$).

To specify a quantum theory, one has to choose a vacuum. Then in order to develop the theory, one has to use the field variables adapted to the corresponding vacuum sector. These field variables must vanish up to a gauge transformation on the gauge orbit corresponding to the chosen vacuum in order that the vacuum expectation values of the associated quantum fields vanish up to a gauge transformation.

Thus the variables A_μ and ϕ_k are adapted to the vacuum sector of Ω_0 corresponding to the gauge orbit of ($A_0 = 0$, $\phi_k = 0$). In this sector, one has an ordinary massless $U(n)$–Yang–Mills field described by the A_μ minimally coupled to the fields ϕ_k, which are massive with the same mass $\mathbf{m}_\phi = n\mathbf{m}$.

The variables adapted to the vacuum sector of Ω_1 are the A_μ and the $\psi_k = \phi_k - i\mathbf{m}E_k$. The translation $\phi_k \mapsto \psi_k$ gives a quadratic term in the traceless part [i.e., the $SU(n)$ part] of the A_μ, which becomes massive with the mass $\mathbf{m}_A = \sqrt{2}n\mathbf{m}$. The $U(1)$ part of the A_μ remains massless and the mass spectrum of the ψ_k becomes complicated. We shall describe this spectrum in the case $n=2$.

D. The case $n=2$ in the sector of Ω_1

The vacuum Ω_1 corresponds to the gauge orbit of pure gauge connections on the free Hermitian $\mathscr{A}$ module of rank 1. The vacuum sector of Ω_1 is therefore very natural from the point of view of the underlying noncommutative differential geometry. We now assume that $\mathscr{A} = C^\infty(\mathbb{R}^{s+1}) \otimes M_2(\mathbb{C})$ and we compute the mass spectrum of the ψ_k (for Ω_1). For that we write ψ_k as $\psi_k = i(\psi_k^0 1 + \psi_k^l E_l)$ and decompose ψ_k^l into its irreducible parts as $\psi_k^l = \tau\delta_k^l + \sigma_k^l + \alpha_k^l$, where $\tau = \frac{1}{3}\psi_l^l$, (σ_k^l) is symmetric and traceless, and (α_k^l) is antisymmetric. One then obtains from (9) and $\psi_k = \phi_k - i\mathbf{m}E_k$ the following mass spectrum. The fields ψ_k^0 have mass $\mathbf{m}_0 = 2\mathbf{m}$, the field τ has mass $\mathbf{m}_\tau = 2\mathbf{m}$, the fields σ_k^l have mass $\mathbf{m}_\sigma = 4\mathbf{m}$, and the fields α_k^l are massless $\mathbf{m}_\alpha = 0$. Notice that, in contrast to the ϕ_k, the ψ_k transform inhomogeneously under a gauge transformation and that one can fix the gauge by imposing $\alpha_k^l = 0$.

E. Generalization

One can generalize similarly the $U(r)$–Yang–Mills action by writing the action for a Hermitian connection on the free Hermitian $\mathscr{A}$ module of rank r. The action again has the form (9) but now the A_μ and the ϕ_k are $(nr \times nr)$-anti-Hermitian-matrix-valued. Thus, using Theorem 4.1, there are as many gauge orbits of connections on which the action vanishes as there are unitary classes of anti-Hermitian representations of $SU(n)$ in $\mathbb{C}^{nr}$. One thus has vacua Ω_α, $\alpha \in \{0,1,\ldots,N(n,r)\}$, for the quantum theory. The number $N(n,r)$ grows very quickly with r for $n \geqslant 2$.

VI. CONCLUSION

For $A = C^\infty(\mathbb{R}^4) \otimes M_n(\mathbb{C})$, i.e., on four-dimensional space-time with $n=2$, the theory described in Secs. V B–V D has similarities with the bosonic part of the standard model of electroweak theory. The ϕ_k plays the role of the Higgs fields and the sector of Ω_1 is similar to the broken phase. One has then a $U(1) \times SU(2)$ gauge theory and the mechanism that produces a mass for the $SU(2)$ part of the gauge field is very similar to the Higgs mechanism. There are, however, two main differences. The first one is that here one has two stable gauge invariant vacua. The second one is that since the ϕ_k or the ψ_k are the components of a Hermitian connection, they are anti-Hermitian and thus they do not interact with the electromagnetic field, i.e., with the $U(1)$ part of the A_μ. Thus there is nothing here like the Weinberg angle and the $U(1)$-gauge field is completely decoupled.

From the point of view of perturbation theory in $\mathbb{R}^4$, the theory we have presented is renormalizable. To carry out the renormalization program one has to use standard BRS technique. However, the usual BRS invariance does not forbid terms like $\mathrm{Tr}(\phi_k^2)$ with arbitrary coefficients. These would break the form of (9), which is the square norm of a curvature and one must therefore find an extended BRS or some other invariance that takes into account the fact that the action is a functional of a curvature. Another point we did not discuss here is the theory of spinor fields in the context of our model. Work on these points is currently in progress.

In Ref. 10, we give an informal discussion of the models of gauge theory presented here with a presentation of the analog of the scalar field for $\mathscr{A} = C^\infty(\mathbb{R}^{s+1}) \otimes M_n(\mathbb{C})$ and a discussion of the analogies and the differences of our work with the theories of Kaluza–Klein type.

[1] A. Connes, "C*-algèbres et géométrie différentielle," C. R. Acad. Sci. Paris Ser. A **290**, 599 (1980).
[2] A. Connes, "Non-commutative differential geometry," Publ. I. H. E. S. **62**, 257 (1986).
[3] M. Karoubi, "Homologie cyclique des groupes et algèbres," C. R. Acad. Sci. Paris Ser. I **297**, 381 (1985).
[4] M. Dubois-Violette, "Dérivations et calcul différentiel noncommutatif," C. R. Acad. Sci. Paris Ser. I **307**, 403 (1988).
[5] M. Dubois-Violette, R. Kerner, and J. Madore, "Non-commutative differential geometry of matrix algebras," preprint ORSAY-LPTHE 88/47, 1988.
[6] J. Madore, "Non-commutative geometry and the spinning particle," Lec-

ture given at Les Journées Relativistes, Geneva, April 1988 and at the XI Warsaw Symposium on Elementary Particle Physics, Kazimierz, May 1988.
[7]C. Itzykson and J. B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980).
[8]A. J. Macfarlane, A. Sudbery, and P. H. Weisz, "On Gell-Mann's λ-matrices, d- and f-tensors, octets, and parametrization of SU(3), Commun. Math. Phys. **11**, 77 (1968).
[9]R. Kerner, L. Nikolova, and V. Rizov, "A two-level Kaluza–Klein theory," Lett. Math. Phys. **14**, 333 (1987).
[10]M. Dubois-Violette, J. Madore, and R. Kerner, "Classical bosons in a noncommutative geometry," preprint ORSAY-LPTHE 88/57, 1988.

On bi-Hamiltonian structures

Jörg Frauendiener and Ezra T. Newman
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 10 August 1989; accepted for publication 20 September 1989)

The definition of a bi-Hamiltonian structure is reviewed, and it is shown that for systems of differential equations of the form $\dot{x} = v(x)$ on even-dimensional manifolds, there always exists locally a bi-Hamiltonian structure. If this structure is "global," then the system of equations is integrable. Furthermore, the geometry and canonical forms for such structures are discussed.

I. INTRODUCTION

In recent years there has been a great deal of interest in systems of first-order differential equations that possess what is referred to as a bi-Hamiltonian structure.[1–6] Most often this interest has centered on the properties of particular systems, with the emphasis frequently being on partial differential equations, e.g., KdV, nonlinear Schrödinger, etc. Due to this interest in particular equations and the inherent complexity of partial differential equations (PDE), as compared to systems of ordinary differential equations (ODE), frequently sight was lost of the general geometric ideas underlying these structures. It is the purpose of this paper to show that there is a very simple geometric idea associated with these bi-Hamiltonian structures, and that in fact every set of equations of the form $\dot{x} = v(x)$ on an even-dimensional manifold possesses, locally, a bi-Hamiltonian structure. When this structure can be extended "globally," then the equations form an "integrable" system. We give a canonical form for a bi-Hamiltonian structure and its related recursion operator, and show the existence of a unique polarization of the phase space of a system that admits a bi-Hamiltonian structure.

In Sec. II we will give some examples and discuss some relevant simple facts and ideas concerning dynamical systems, i.e., systems of equations of the form

$$\dot{x}^a = v^a(x), \quad a = (1,2,...,N). \tag{1.1}$$

(We note that many PDE's can be written in this form if we let N go to infinity. Questions of convergence, existence, etc., in the infinite-dimensional case, will not concern us.)

In Sec. III we will give a brief outline of symplectic geometry and define what is meant by a bi-Hamiltonian structure. We furthermore discuss properties of bi-Hamiltonian systems. Though this discussion is a combination of older[1,6] and new material we present it in a unified manner. Roughly speaking, a bi-Hamiltonian structure associated with a system (1.1) consists of two independent symplectic structures (with an important compatibility condition on them) and two different Hamiltonians, so that (1.1) gives the canonical equations of motion for both Hamiltonians. In Sec. IV we will show that every system of the form (1.1) can be given, in a trivial fashion, locally, a bi-Hamiltonian structure, and that in fact many such bi-Hamiltonian structures exist. One can furthermore see, in this simple case, certain general properties of bi-Hamiltonian structures. We will also show that if one begins with a Hamiltonian system, i.e., where (1.1) is given as a Hamiltonian vector field, then there will exist, again locally, an alternative Hamiltonian and alternative symplectic form with the required properties, so that it also becomes a bi-Hamiltonian system. We will see that when this structure is global, the equations are those of an integrable system.

II. LOCAL DYNAMICAL SYSTEMS

We will consider a vector field $v^a(x)$ on an even-dimensional manifold, dim $= 2N$, with local coordinates x^a. We will restrict ourselves to an open region u, a neighborhood of a point where v^a does not vanish. u will be taken sufficiently small so that it is foliated by a $(2N-1)$-dimensional set of integral curves of v^a, i.e., the curves crossing some $(2N-1)$-dimensional "initial data surface." These integral curves satisfy the set of equations $\dot{x}^a = v^a(x)$, $a = (1,2,...,2N)$. Here are some examples.

(1) Damped harmonic oscillator:

$$\ddot{x} + a\dot{x} + bx = 0,$$

or

$$\dot{x}^1 = x^2, \quad \dot{x}^2 = -ax^2 - bx^1.$$

(2) Canonical equations of motion:

$$x^a = (q^i, p_i), \; v^a = \left(\frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial q^i}\right).$$

(3) Nonlinear partial differential equations:

$$\dot{u} = F(u, u_y, u_{yy}, ...),$$

where $u = u(t,y)$, y is a point in some finite-dimensional manifold M, and F is a nonlinear function of its arguments. If $u_a(y)$ is a orthonormal basis on M, then one can write

$$u(t,y) = \sum_a x^a(t) u_a(y),$$

from which it follows, at least formally, that $\dot{x}^a = v^a$, $a = (1,2,...,\infty)$, with

$$v^a = \int u_a(y) F \, dy.$$

We will not concern ourselves with the issue of convergence or existence of this representation of the equation $\dot{u} = F(u, u_y, u_{yy}, ...)$. We note, however, that most of the ideas expressed here can be reformulated and restated in the language of function spaces, permitting a complete discussion of partial differential equations. Our representation is intended only to give a uniform discussion of the geometric

 0022-2488/90/020331-07$03.00

ideas common to both partial and ordinary differential equations.

By assumption, there is a $(2N-1)$-dimensional surface in U which can be given local coordinates y^i, so that the integral curves have the form

$$x^a = f^a(y^i,t), \quad i = (2,3,...,2N). \tag{2.1}$$

If we now consider a change of local coordinates $x^a \mapsto y^a = (t := y^1, y^i)$ given by (2.1), then (1.1) becomes, in the new (special) coordinates,

$$\dot{y}^a = \delta^a_1. \tag{2.2}$$

In other words, locally, all first-order systems (1.1) become trivial and equivalent. In practice this transformation depends on being able to integrate the original equations and is thus usually not of any value; however, in principle, it makes proving certain local theorems very simple. We will use this result frequently.

The *local symmetries* of a system (1.1) can be defined in the following way: if we are given an arbitrary single integral curve of (1.1), say $x^a = x^a(t)$, and a vector field $f^a(x)$ such that

$$x^a = x^a(t) + \epsilon f^a(x(t)) \tag{2.3}$$

is also an integral curve, i.e., also a solution of (1.1), then we say that $f^a(x)$ is a local symmetry. Substituting (2.3) into (1.1) we obtain the condition on $f^a(x)$ that is Lie derivative along v vanishes, i.e.,

$$\mathcal{L}_v f^a \equiv [v,f]^a \equiv v^b \partial_b f^a - f^b \partial_b v^a = 0. \tag{2.4}$$

One now has the natural question, given the field v^a, how many local symmetries or fields f^a, other than v itself, exist? It is clear from (2.2) that the answer is $2N-1$. In the special coordinate system of (2.2) we have the $2N-1$ independent solutions of (2.4), with $i = (2,3,...,2N)$:

$$f^a_{(i)} = \delta^a_i. \tag{2.5}$$

Note that the $f_{(i)}$ also have vanishing Lie derivatives among themselves. This means that if we include the original v as one of the f's, i.e., $v := f_{(1)}$, then the original differential equation determines (not uniquely) $2N$ vector fields with vanishing Lie derivatives among themselves on U. Among the set of vectors $f^a_{(i)}$ we will refer to $v = f_{(1)}$ as the *primary field*.

Note further that these $2N$ f's define a stepping operator $\Sigma^a{}_b$, so that

$$\Sigma^a{}_b f^b_{(i)} = f^a_{(i+1)}. \tag{2.6}$$

If one defines the dual basis $f^{(i)}_b$ to the vectors $f^b_{(i)}$, i.e., by $f^{(i)}_b f^a_{(i)} = \delta^a_b$, then $\Sigma^a{}_b$ can be defined by

$$\Sigma^a{}_b := \sum_{i=1}^{2N} f^a_{(i+1)} f^{(i)}_b, \tag{2.7}$$

where we use $f^a_{(2N+1)} \equiv f^a_{(1)}$. It is obvious that (2.6) is satisfied.

Equations (2.6) and (2.7) can be generalized in the following fashion:

$$\Sigma^a{}_b := \sum_{i=1}^{2N} f^a_{\pi(i)} f^{(i)}_b, \tag{2.8}$$

where π is some permutation of the indices $i = 1,2,3,...,2N$; Eq. (2.6) is then a special case involving the cyclic permutation on the $2N$ values of i. Using the special coordinate system, $\Sigma^a{}_b$ would have the form of a permutation of the rows (or columns) of the unit matrix and is thus clearly nonsingular. Note that the action of Σ defined by (2.8) is slightly different from that defined by (2.7): in (2.7) repeated action of Σ on any vector $f^b_{(i)}$ cycles it among all $2N$ vectors, while for (2.8) it might get cycled over a lower rank cycle, and then other starting vectors are needed to obtain all $2N$ vectors.

Of particular interest to us is the case where the set of $2N$ vectors is divided into two sets of N each with some N cycle taking place over each set. This choice of Σ is a special case of what we will refer to as a *recursion operator*, denoted by S; it will play a crucial role later.

III. SYMPLECTIC AND BI-HAMILTONIAN STRUCTURES

A. Symplectic structure

A symplectic structure on U is a tensor field Ω_{ab} satisfying the following conditions: (1) It is skew; (2) it is nondegenerate; and (3) it is closed, i.e., $\partial_{[a}\Omega_{bc]} = 0$.

We will impose a further condition on Ω, namely, that its Lie derivative along some vector field v^a (the primary vector field of Sec. II) must vanish, i.e.,

$$\mathcal{L}_v \Omega_{ab} = 0. \tag{3.1}$$

The importance of this latter condition is that it allows the primary vector field to be considered (as is always possible, e.g., Ref. 7) as a Hamiltonian vector field, i.e., to be derivable from some Hamiltonian $H(x^a)$ via

$$v^a = \Omega^{ab} \partial_b H, \tag{3.2}$$

where Ω^{ab} is the inverse of Ω_{ab}, i.e., $\Omega_{ab}\Omega^{bc} = \delta_a{}^c$. The equations (1.1) would then be the canonical equations of motion with Hamiltonian H.

These conditions can easily be satisfied locally, and in fact there are many inequivalent symplectic structures satisfying (3.1). To see this one takes the $2N$ dual forms $f^{(i)}_a$ and arbitrarily separates them into two groups of N, each indexed respectively by (α) and $(N+\alpha)$, $\alpha = 1,2,...,N$, and thereby paired via the same value of α, so that we have $f^{(i)}_a = \{f^{(\alpha)}_a, f^{(\alpha+N)}_a\}$.

We can define

$$\Omega_{ab} := \sum_{\alpha=1}^{N} f^{(\alpha)}_{[a} f^{(\alpha+N)}_{b]}. \tag{3.3}$$

From the construction it is clear that conditions (1) and (2) are satisfied. Condition (3) and Eq. (3.1) are also satisfied, which can be seen by first going to the special coordinates (2.5) where they are immediately satisfied; hence they are true in any coordinate system since they are tensor equations.

The inverse to Ω_{ab} so defined, namely, Ω^{ab}, is easily seen to be given by

$$\Omega^{ab} := \sum_{\alpha=1}^{N} f^{[a}_{(\alpha)} f^{b]}_{(\alpha+N)}. \tag{3.4}$$

B. Bi-Hamiltonian structure

A bi-Hamiltonian structure on U is defined as two symplectic structures Ω_{ab} and $\tilde{\Omega}_{ab}$ on U that satisfy the following relationship: construct the (1,1) tensor, referred to as the recursion operator,

$$S^b{}_a = \Omega^{bc}\tilde{\Omega}_{ca}, \tag{3.5}$$

and require that its associated Nijenhuis tensor (or torsion tensor) N_S—or simply N—vanish, i.e.,

$$N^c_{ab} := 2S^d{}_{[a}S^c{}_{b],d} + 2S^c{}_dS^d{}_{[a,b]} = 0, \tag{3.6}$$

or abstractly

$$N(X,Y) = [SX,SY] + S^2[X,Y] - S[SX,Y] - S[X,SY] = 0 \tag{3.6a}$$

for any vectors X and Y. Equation (3.6a) is more useful and easier to manipulate than (3.6).

If furthermore the Lie derivatives of both Ω_{ab} and $\tilde{\Omega}_{ab}$ along the primary vector field v^a vanish, i.e.,

$$\mathscr{L}_v\Omega_{ab} = 0 \quad \text{and} \quad \mathscr{L}_v\tilde{\Omega}_{ab} = 0 \tag{3.7}$$

are also satisfied, then we will say that it is the bi-Hamiltonian structure associated with v^a.

The proof of the following two theorems on the algebraic implications of the existence of a bi-Hamiltonian structure are given in Appendix A.

Theorem 1: Recursion operators [defined by (3.5)] have the following algebraic property: given any vector v in the $2N$-dimensional tangent space, then the set of $N+1$ vectors

$$\{v,Sv,S^2v,S^3v,S^4v,...,S^Nv\} \tag{3.8}$$

is linearly dependent.

Note that there is no implication that the first N vectors in (3.8) are linearly independent; one can easily construct cases where they are and cases where they are not. In the case where there are N linearly independent vectors we will refer to the bi-Hamiltonian structure as a *maximal bi-Hamiltonian structure.*

Theorem 2: Given a maximal bi-Hamiltonian structure Ω, $\tilde{\Omega}$, then at each point there exists a basis of the tangent space at that point such that the matrix representation of Ω, $\tilde{\Omega}$, and S are

$$\Omega = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \tilde{\Omega} = \begin{pmatrix} 0 & A^t \\ -A & 0 \end{pmatrix}, \quad S = \Omega^{-1}\tilde{\Omega} = \begin{pmatrix} A & 0 \\ 0 & A^t \end{pmatrix},$$

with

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & a_0 \\ 1 & 0 & \cdots & 0 & a_1 \\ 0 & 1 & \cdots & 0 & a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_{n-1} \end{pmatrix}.$$

The invariance group of this structure is an N-dimensional subgroup of $\mathrm{Sp}(N)$, the symplectic group in $2N$ dimensions, that is isomorphic to $\mathbf{R}^N$.

We note that from the skew symmetry of Ω and $\tilde{\Omega}$, the characteristic polynomial of S is a perfect square, i.e., that

$$p_S(\lambda) = \det(S^a{}_b - \lambda\delta^a{}_b) = p_A(\lambda)^2 = (\lambda^N - a_{N-1}\lambda^{N-1} - a_{N-2}\lambda^{N-2} - \cdots - a_1\lambda - a_0)^2.$$

This shows that every eigenvalue is degenerate with an even multiplicity. Note that this does not imply that S is diagonalizable. It is easy to construct examples where it is not. If, for a maximal bi-Hamiltonian structure, the recursion operator is such that

$$v = S^Nv,$$

then we will refer to the bi-Hamiltonian structure as a perfect bi-Hamiltonian structure and to the recursion operator as a perfect recursion operator.

If we try to extend the algebraic normal form of S from a point to a local neighborhood of that point, we need integrability conditions. Here the Nijenhuis condition comes into play.

Remark 1: We point out the following properties and equivalent formulations of the Nijenhuis condition. Since the proofs are either straightforward calculations or rather technical we omit them.

(i) If S can be written as in (3.5), we have the following equivalence: $N_S = 0$ iff for every closed one-form ω_a the one-form $S^a{}_b\omega_a$ is also closed. This is an important result, since it implies that S "maps gradients into gradients." For general (1,1) tensor fields, $N_S = 0$ is not sufficient for this property.

(ii) For arbitrary real numbers λ, the (2,0) tensor field $\Omega^{ab} + \lambda\tilde{\Omega}^{ab}$ is the inverse of a symplectic form iff $N_S = 0$. This calculation is rather lengthy.

(iii) If S satisfies the Nijenhuis condition, so does its inverse and all its powers.

Theorem 3: Assume that we have a bi-Hamiltonian structure associated with the primary vector field v^a. It then follows that the vectors

$$f_{(1)} = v, \quad f_{(2)} = Sv, \quad f_{(3)} = S^2v,... \tag{3.9}$$

have a vanishing Lie derivative with each other, i.e.,

$$\mathscr{L}_{f(\alpha)}f_{(\beta)} = [f_{(\alpha)}, f_{(\beta)}] = 0, \quad \alpha,\beta = 1,2,... . \tag{3.10}$$

If the bi-Hamiltonian structure is maximal for v^a, then the first N vectors form a linearly independent set of symmetries of v^a.

Proof: We give an outline of the proof here; the complete proof is given in Appendix A. We first note that from (3.7) we have $\mathscr{L}_vS = 0$, which implies that v commutes with the other f's:

$$\mathscr{L}_v(Sv) = [v,Sv] = 0, \quad \mathscr{L}_v(S^2v) = [v,S^2v] = 0,... . \tag{3.11}$$

This result is used as the first step in an induction process on the number of S's appearing in the commutator. The induction step is carried out by using the Nijenhuis condition (3.6). ■

We see that the main result of requiring the vanishing of the torsion tensor is to make the set of vectors $f^a_{(\alpha)}$ [i.e., (3.9)] commuting symmetries of v^a. Another consequence of this theorem is as follows: From the algebraic results of Appendix A we know that in the maximal case at every point of the manifold there exists a Lagrangian subspace generated by v and its S iterates. Theorem 3 then tells us that the distribution so defined is locally integrable, i.e., that the phase space can be locally foliated by Lagrangian submanifolds. In other words, there exists a preferred polarization.

In the case where we have a maximal bi-Hamiltonian structure, we can write $f^a_{(N+1)} = a_0 f^a_{(1)} + a_1 f^a_{(2)} + \cdots + a_{N-1} f^a_{(N)}$, where $a_0,\ldots,a_{N-1}$ are now functions on the phase space. Not surprisingly, they are constant on the Lagrange submanifolds; namely, we have

$$\begin{aligned} 0 &= [f_{(k)}, f_{(N+1)}] \\ &= da_0(f_{(k)}) f_{(1)} \\ &\quad + da_1(f_{(k)}) f_{(2)} + \cdots + da_{N-1}(f_{(k)}) f_{(N)}. \end{aligned}$$

Since $f_{(1)},\ldots,f_{(N)}$ are linearly independent at every point we obtain

$$\mathscr{L}_{f_{(\alpha)}} a_k = 0, \quad \text{for } 0 \leqslant \alpha, k \leqslant N-1.$$

Hence the coefficients of the minimal polynomial of a maximal bi-Hamiltonian structure are conserved quantities. Note that in the case of a perfect S we have $a_0 = 1$, and all the other a's vanish. In other references [8–11] the emphasis lies on another set of conserved quantities, viz., the traces of the various powers of S. Clearly, it is possible to express one set in terms of the other. However, this formula is quite complicated and will not be given here. It is a common misbelief that in a bi-Hamiltonian structure the a's have to be nonconstant functions. We will show below that is not necessarily true.

We now come to the main theorem of bi-Hamiltonian structures.[1]

Theorem 4: Given a maximal bi-Hamiltonian structure associated with the primary vector field v^a, there will exist N scalar functions $H_{(\alpha)}$ that are unique up to constants, pairs of which act as the two Hamiltonians (associated with the two different symplectic forms Ω_{ab} and $\widetilde{\Omega}_{ab}$) for each of the N vector fields $f^a_{(\alpha)}$; more specifically, we have

$$f^a_{(\alpha)} = \Omega^{ab}\,\partial_b H_{(\alpha)} = \widetilde{\Omega}^{ab}\,\partial_b \widetilde{H}_{(\alpha)}, \quad \alpha = 1,2,\ldots,N, \tag{3.12}$$

with $\widetilde{H}_{(\alpha)} = H_{(\alpha+1)}$ and $H_{(N+1)}$ a linear combination of the $H_{(\alpha)}$.

Proof: Again, the complete proof is given in Appendix B. We show that all the vector fields are Hamiltonian for both symplectic structures by showing that they Lie derive the symplectic forms. Then we show that there exists a ladder-type relationship among the $H_{(\alpha)}$, i.e., the Hamiltonian $H_{(\alpha)}$ associated with $\widetilde{\Omega}$ becomes the Hamiltonian for the next vector but now associated with the Ω. ■

Remark 2: (a) All the $f^a_{(\alpha)}$ are Hamiltonian vector fields for both Ω_{ab} and $\widetilde{\Omega}_{ab}$, i.e., $\mathscr{L}_f \Omega_{ab} = \mathscr{L}_f \widetilde{\Omega}_{ab} = 0$, for all $f^a_{(\alpha)}$.

(b) For each vector $f^b_{(\alpha)}$ there are (via 3.12) *two* associated one-forms, $\partial_b H_{(\alpha)}$ and $\partial_b H_{(\alpha+1)}$, and, conversely, for each one-form $\partial_b H_{(\alpha)}$ there are two associated vectors, $f^b_{(\alpha)}$ and $f^b_{(\alpha-1)}$. Furthermore, we have that $\widetilde{\Omega}^{ab}\,\partial_b H_{(\alpha+1)} = \Omega^{ab}\,\partial_b H_{(\alpha)}$ and hence $\partial_b H_{(\alpha)}\, S^b{}_a = \partial_b H_{(\alpha+1)}$, i.e., S acts as a recursion operator on the one-forms as well as on vectors (cf. Remark 1).

Theorem 5: The functions H_α of Theorem 4 have vanishing Poisson brackets with each other with respect to either symplectic structure, i.e.,

$$\begin{aligned} \{H_\alpha, H_\beta\} &= \Omega^{ab}\,\partial_a H_{(\alpha)}\,\partial_b H_{(\beta)} = 0, \\ \{H_\alpha, H_\beta\}^\sim &= \widetilde{\Omega}^{ab}\partial_a H_{(\alpha)}\partial_b H_{(\beta)} = 0. \end{aligned} \tag{3.13}$$

Proof: The assertion follows from the fact that the Poisson bracket between two functions is equal to the Lagrange bracket of the corresponding Hamiltonian vector fields, which vanishes by Lemma 1 of Appendix A. ■

As an immediate consequence of Theorem 5 we find that if the bi-Hamiltonian structure is maximal and global then the system is Liouville integrable.

IV. EXISTENCE OF BI-HAMILTONIAN STRUCTURES

In this section we will discuss the existence of bi-Hamiltonian structures in three different contexts. Specifically, we will study (1) the local existence when there is just a dynamical system, i.e., $\dot{x} = v(x)$; (2) the local existence when there already exists a Hamiltonian structure, i.e., $\dot{x} = \Omega^{-1}\nabla H$; and (3) the global existence when there already exists a Hamiltonian structure.

A. Local existence

Theorem 6: Every dynamical system, in the neighborhood of a regular point of the field, possesses a bi-Hamiltonian structure and, in fact, many such structures. In particular, one can always choose the bi-Hamiltonian structure so that it is perfect.

Proof: The proof of the above contention is most easily given by using the special coordinate system of Sec. II. In Eq. (3.3) we saw that we could divide the set of vectors $f^b_{(i)}$ and dual vectors $f^{(i)}_a$ into two groups of N, each indexed respectively by (α) and $(N+\alpha)$, $a = 1,2,\ldots,N$, thereby pairing them via the same value of α, so that we have, e.g., $f^a_{(i)} = \{f^a_{(\alpha)}, f^a_{(N+\alpha)}\}$. We defined the symplectic form in (3.3) by

$$\Omega_{ab} := \sum_{\alpha=1}^{N} f^{(\alpha)}_{[a} f^{(\alpha+N)}_{b]}.$$

In the $f^a_{(i)}$ basis we have that

$$\Omega_{ab} = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix}.$$

A second symplectic form, $\widetilde{\Omega}$, can be constructed from the same set of vectors in the following manner: Consider the set $f^a_{(i)} = \{f^a_{(\alpha)}, f^a_{(N+\alpha)}\}$; now reorder the first N vectors by performing a permutation on the index, $\alpha \mapsto \pi\alpha$, or $f^a_{(\alpha)} \mapsto f^a_{(\pi\alpha)}$; we obtain the new paired set $f^a_{(i)} = \{f^a_{(\pi\alpha)}, f^a_{(N+\alpha)}\}$. We now construct

$$\widetilde{\Omega}_{ab} := \sum_{\alpha=1}^{N} f^{(\pi\alpha)}_{[a} f^{(\alpha+N)}_{b]}, \tag{4.1}$$

or

$$\widetilde{\Omega} = \begin{pmatrix} 0 & A \\ -A^t & 0 \end{pmatrix},$$

where A is the permutation π of the columns of the identity matrix $\mathbb{1}$. We have for the recursion operator $S = \Omega^{-1}\widetilde{\Omega}$ [see Remark 2 (b) at the end of Sec. III]

$$S=\begin{pmatrix} -A^t & 0 \\ 0 & -A \end{pmatrix}.$$

The Nijenhuis condition is obviously satisfied since Ω and $\tilde{\Omega}$ are constant, and we thus have a bi-Hamiltonian structure. It is easily seen that if the permutation is the cycle $(1,2,...,N) \mapsto (N,1,2,...,N-1)$, the bi-Hamiltonian structure and recursion operator are perfect. ■

B. Local existence with a Hamiltonian

We now assume that the primary field v is already a Hamiltonian field, i.e., there exists a symplectic structure Ω and a Hamiltonian $H_{(1)}$. It thus takes the form

$$v^a = \Omega^{ab}\,\partial_b H_{(1)}.$$

It is well known[12] that *locally* there will always exist N functions (including $H_{(1)}$) on the $2N$-dimensional phase space that have vanishing Poisson brackets with each other, i.e., are in involution. We will refer to them as $H_{(\alpha)}$, with $H_{(N+1)} = H_{(1)}$. Locally, they all define Hamiltonian vector fields,

$$f^a_{(\alpha)} = \Omega^{ab}\,\partial_b H_{(\alpha)} \quad \alpha = 1,2,...,N,$$

which in canonical coordinates (q,p) have the form

$$f^a_{(\alpha)} = (\partial_p H_{(\alpha)}, -\partial_q H_{(\alpha)}).$$

We can now consider a local canonical transformation whose form is

$$P_{(\alpha)} = H_{(\alpha)}(q,p),$$
$$Q^{(\alpha)} = Q^{(\alpha)}(q,p).$$

The existence of a generating function for this type of canonical transformation, i.e., where the H's become the new p's, is guaranteed by the involutive property of the H's. In this new canonical coordinate system we have that

$$f^a_{(\alpha)} = (\delta^a_{(\alpha)},0).$$

By defining $\tilde{H}_{(\alpha)} = H_{(\alpha+1)}$ and using (3.12),

$$f^a_{(\alpha)} = \Omega^{ab}\,\partial_b H_{(\alpha)} = \tilde{\Omega}^{ab}\,\partial_b \tilde{H}_{(\alpha)}, \quad \alpha = 1,2,...,N,$$

we obtain in the new canonical coordinate system that

$$\Omega_{ab} = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix} \quad \text{and} \quad \tilde{\Omega} = \begin{pmatrix} 0 & A \\ -A^t & 0 \end{pmatrix},$$

with A being the first permutation on the columns of the identity matrix.

We thus have the result that locally every Hamiltonian system can be extended into a perfect bi-Hamiltonian system.

C. Global bi-Hamiltonians

In the above local proof of the bi-Hamiltonian structure we assumed that we knew N local involutive functions $H_{(\alpha)}$ which always exist. If, however, we had assumed that these functions were global (true only in special systems known as integrable systems), then the proof of the bi-Hamiltonian structure would have been exactly the same, and we would have had a global bi-Hamiltonian structure, i.e., an integrable system implies a global perfect bi-Hamiltonian structure.

The converse, that a global perfect bi-Hamiltonian structure implies an integrable system, follows from the global existence of the N involutive independent $H_{(\alpha)}$'s in Eq. (3.12). ■

Note that the bi-Hamiltonian structures we have constructed are in some sense trivial; they do not contain any information about the dynamical system that we are considering. Also, it is worthwhile to point out that the bi-Hamiltonian structure of a system is by no means uniquely determined.

V. DISCUSSION

We have shown that all dynamical systems on an even-dimensional manifold locally possess many bi-Hamiltonian structures. However, for specific equations it will, in general, be extremely difficult or impossible to construct even a single one. Even in the case of a Hamiltonian system the problem of finding a second symplectic form usually depends on finding N involutive (local) integrals of the system. Bi-Hamiltonian systems thus do not seem to be of great practical use for general systems, becoming of use only when the system is integrable. It nevertheless appears that the existence of such a structure might be of use in theoretical discussions of dynamical systems. In a future paper this possibility will be investigated in connection with the existence of Lax pairs (i.e., linear equations whose integrability conditions yield the dynamical system in question) for general dynamical systems.

Bi-Hamiltonian might also apply to the problem of the quantization of general dynamical systems. The usual quantization procedure suffers from the fact that there is no unique polarization of the phase space that would allow one to uniquely reduce the number of independent variables in the wave functions from $2N$ to N. If, however, there is a bi-Hamiltonian structure for the system in question, then there also exists a preferred polarization. Those situations where one usually believes that the quantization process is understood appears to be either a integrable system or perturbations off an integrable system.

ACKNOWLEDGMENTS

J. F. would like to thank the Alexander von Humboldt Foundation for financial support by a Feodor–Lynen Fellowship during his stay in Pittsburgh, and the Department of Physics and Astronomy of the University of Pittsburgh for kind hospitality.

E. T. N. thanks the National Science Foundation for support on Grant No. PHY 80023.

APPENDIX A: ALGEBRAIC PROPERTIES OF A BI-HAMILTONIAN STRUCTURE

Let E be a real vector space of dimension $2N$. Let Ω and $\tilde{\Omega}$ be skew nonsingular bilinear forms on E. We define an isomorphism of E by $S^b{}_a := \Omega^{bc}\tilde{\Omega}_{ca}$ or, equivalently, by

$$\tilde{\Omega}(u,v) = \Omega(Su,v), \quad \text{for all } u,v \text{ in } E. \tag{A1}$$

The fact that $\tilde{\Omega}$ is skew implies

$$\Omega(Su,v) = \Omega(u,Sv), \tag{A2}$$

for arbitrary vectors u,v in E. Note that there is a 1–1 correspondence between skew nonsingular bilinear forms and linear isomorphisms that satisfy condition (A2). We will focus on S most of the time and then obtain statements about $\tilde{\Omega}$ by means of (A1).

A first simple consequence of the definition of S is the following: the characteristic polynomial of S is given by

$$p_s(\lambda) = \det(S^a{}_b - \lambda\delta^a{}_b) = \det(\Omega^{ab})\det(\tilde{\Omega}_{ab} - \lambda\Omega_{ab}).$$

Since Ω_{ab} and $\tilde{\Omega}_{ab}$ are skew, and since the determinant of a skew matrix is a perfect square, the square of its Pfaffian,[13] we obtain $p_S(\lambda) = (q_N(\lambda))^2$, where q_N is a polynomial in λ of degree N. This implies that the roots of p_S (as a polynomial over the complex field) come in pairs—the multiplicity of each root is even. From the Hamilton–Cayley theorem we know that S annihilates its characteristic polynomial. If S is diagonalizable, then it also must annihilate q_N. To show that this is true in general, we move on to give a canonical form for the matrix of S.

Lemma 1: Given Ω and S satisfying (A2), we have (i) $\Omega(S^n v, S^m v) = 0$, for v in E and arbitrary integers n, m; and (ii) For every vector v in E, the set of $N+1$ vectors $\{v, Sv, S^2 v, ..., S^N v\}$ is linearly dependent.

Proof: Note that $\Omega(S^n v, S^n v) = 0$, and that from (A2), $\Omega(S^n v, S^{n+1} v) = \Omega(S^{n+1} v, \quad S^n v) = -\Omega(S^n v, S^{n+1} v) = 0$. In the general case we can use (A2) to put $\Omega(S^n v, S^m v)$ into one of the above forms. This proves (i). Point (ii) follows from the fact that Ω restricted to the span of the iterates of v vanishes. If this were an $(N+1)$-dimensional subspace, this would be in contradiction with the regularity of Ω. ■

For every v, there is a maximal number m such that the first m iterates of v are linearly independent. By Lemma 1, $m \leqslant N$. If v is such that $m = N$, we say that *S is maximal for v*. In the following we will restrict ourselves to the case where S is maximal for some v, since this is the most important case for us and in the applications.

Let $v_i := S^i v$. Then $\Omega(v_i, v_k) = 0$ and $S^N v = a_0 v + a_1 v_1 + \cdots + a_{N-1} v_{N-1}$, for N real numbers $a_0, a_1, ..., a_{N-1}$. From a general result of linear algebra[13] we can put Ω in Darboux form, i.e., we can find N vectors $u_0, u_1, ..., u_{N-1}$ such that the set $\{v_0, v_1, ..., v_{N-1}, u_0, u_1, ..., u_{N-1}\}$ constitutes a basis of E, and such that $\Omega(u_i, u_k) = \Omega(v_i, v_k) = 0$ and $\Omega(v_i, u_k) = \delta_{ik}$. From (A1) we also have $\tilde{\Omega}(v_i, v_k) = 0$. In this basis the matrix representation of Ω, $\tilde{\Omega}$, and S are the following:

$$\Omega = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix}, \quad \tilde{\Omega} = \begin{pmatrix} 0 & A^t \\ -A & B \end{pmatrix},$$

$$S = \Omega^{-1}\tilde{\Omega} = \begin{pmatrix} A & -B \\ 0 & A^t \end{pmatrix},$$

where B is an arbitrary skew $(N \times N)$ matrix. We know that the matrix of S restricted to the subspace spanned by v and its iterates is

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & a_0 \\ 1 & 0 & \cdots & 0 & a_1 \\ 0 & 1 & \cdots & 0 & a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_{N-1} \end{pmatrix}.$$

We are now in a position to prove Theorem 2 from Sec. III.

Theorem 2: Given a maximal bi-Hamiltonian structure, Ω, $\tilde{\Omega}$ on E, there exists a basis such that the matrix representations of Ω, $\tilde{\Omega}$, and S are

$$\Omega = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix}, \quad \tilde{\Omega} = \begin{pmatrix} 0 & A^t \\ -A & 0 \end{pmatrix}, \quad S = \Omega^{-1}\tilde{\Omega} = \begin{pmatrix} A & 0 \\ 0 & A^t \end{pmatrix},$$

with A as above. The invariance group of this structure is an N-dimensional subgroup of $\mathrm{Sp}(N)$, the symplectic group in $2N$ dimensions, that is isomorphic to $\mathbb{R}^N$.

Proof: We try to find symplectic transformations with respect to Ω that transform B in $\tilde{\Omega}$ to 0. We consider the ansatz

$$T = \begin{pmatrix} 1 & \alpha \\ 0 & \beta \end{pmatrix}.$$

For T to be symplectic for Ω, it is necessary that $\beta = 1$ and that α is symmetric. Then $\tilde{\Omega}$ is transformed into

$$T^t \tilde{\Omega} T = \begin{pmatrix} 0 & A^t \\ -A & B - A\alpha + \alpha A^t \end{pmatrix}.$$

Now T will transform $\tilde{\Omega}$ into the desired form if we can find a solution α to the equation

$$B = A\alpha - \alpha A^t, \tag{A3}$$

for arbitrary skew B. Consider the linear map $\alpha \mapsto A\alpha - \alpha A^t$, which maps the $\frac{1}{2}N(N+1)$-dimensional vector space of symmetric matrices into the $\frac{1}{2}N(N-1)$-dimensional space of skew matrices. We will show that the null space of this map is exactly N-dimensional, which implies by linearity that the map is onto and hence there exists a solution for every skew B. In fact, the solution space will be N-dimensional, which implies that the invariance group is an N-dimensional subgroup of the symplectic group $\mathrm{Sp}(N)$. It is Abelian, and each of its elements is uniquely determined by N real numbers (the last row of the corresponding α). This shows that it is isomorphic to $\mathbb{R}^N$.

In order to solve the homogeneous equation (A3), write α in the form

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix},$$

with N row vectors $\alpha_1, ..., \alpha_N$. The product $A\alpha$ can then be written as

$$A\alpha = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_N \end{pmatrix} = \begin{pmatrix} a_0\alpha_N \\ a_1\alpha_N + \alpha_1 \\ a_2\alpha_N + \alpha_2 \\ \vdots \\ a_{N-1}\alpha_N + \alpha_{N-1} \end{pmatrix}.$$

The homogeneous equation (A3) then requires $A\alpha$ to be symmetric. In terms of its matrix elements this reads $\beta_{ik} = a_{i-1}\alpha_{Nk} + \alpha_{i-1\,k} = a_{k-1}\alpha_{Ni} + \alpha_{k-1\,i} = \beta_{ki}$. From this equation we can determine the matrix elements of α in the following way: we specify freely the last row $\alpha_{N1}, ..., \alpha_{NN}$, and then use the equation as a recursion relation to determine the k th row from the previously known $(k-1)$th row, keeping in mind that α is symmetric. The first row is given by $\alpha_{1\,i-1} = a_0\alpha_{Ni} - a_{i-1}\alpha_{N1}$, for $1 \leqslant i \leqslant N$, $\alpha_{1N} = \alpha_{N1}$. For the

kth row we have $\alpha_{ki-1} = \alpha_{k-1i} + (a_{k-1}\alpha_{Ni} - a_{i-1}\alpha_{Nk})$, for $(k+1) \leqslant i \leqslant N$, $\alpha_{kN} = \alpha_{Nk}$. This completes the proof. ■

This result shows again that the characteristic polynomial of S is $p_S(\lambda) = (p_A(\lambda))^2$, a perfect square. Also, this implies that S annihilates the Nth degree polynomial $p_A(\lambda)$. Now, the minimal polynomial of S, by definition, divides p_A. On the other hand, the assumption of maximality of S implies that its minimal polynomial is at least of degree N, hence it must be equal to p_A. A simple calculation shows that

$$p_A(\lambda) = (-1)^N(\lambda^N - a_{N-1}\lambda^{N-1} - a_{N-2}\lambda^{N-2} - \cdots - a_1\lambda - a_0).$$

We have proved the following corollary.

Corollary: A maximal bi-Hamiltonian structure on a $2N$-dimensional vector space is completely characterized by its *invariant* $(a_0, a_1, ..., a_{N-1})$, an N-tuple of real numbers.

The geometry of the situation is as follows: Given S and v such that S is maximal for v, there exists a unique Lagrangian subspace L for *both* symplectic forms Ω and $\widetilde{\Omega}$ that contains v. The remaining $\mathbb{R}^N$ freedom corresponds to the choice of different transversal Lagrangian subspaces. It is effectively the choice of one vector transversal to L. Then S determines a unique Lagrangian subspace containing this vector that is transverse to L.

APPENDIX B: PROOF OF THEOREMS 3 AND 4

Theorem 3: Assume that we have a bi-Hamiltonian structure associated with the primary vector field v^a. It then follows that the vectors (3.9),

$$f_{(1)} = v, \quad f_{(2)} = Sv, \quad f_{(3)} = S^2 v, ...$$

have a vanishing Lie derivative with each other, i.e.,

$$\mathscr{L}_{f_{(\alpha)}} f_{(\beta)} = [f_{(\alpha)}, f_{(\beta)}] = 0, \quad \alpha, \beta = 1,2,... .$$

If the bi-Hamiltonian structure is maximal for v^a, then the first N vectors form a linearly independent set of symmetries of v^a.

Proof: The first statement is proved in the following manner: We first note that from (3.7) we have $\mathscr{L}_v S = 0$. This can be written as

$$[v, SX] = S[v, X], \tag{B1}$$

for every vector field X. Hence we have $[v, S^{n+1}v] = S[v, S^n v]$. Using induction on n—the case $n = 0$ being trivial—we find that v commutes with all $S^m v$ for all positive integers m.

We now show that $[S^{n-m}v, S^m v] = 0$, for all $0 \leqslant m \leqslant n$, $n > 1$. We use induction on n. The case $n = 2$ is proved with the above result. Now suppose the assertion is true for all $0 \leqslant m \leqslant k$, $k \leqslant n$. The Nijenhuis condition (3.6a) can be written in the form

$$[SX, SY] = S[SX, Y] + S[X, SY] - S^2[X, Y],$$

for arbitrary vectors X, Y. From this we obtain

$$[S^{n+1-m}v, S^m v] = S[S^{m+1-m}v, S^{m-1}v] + S[S^{n-m}v, S^m v] - S^2[S^{n-m}v, S^{m-1}v],$$

but this vanishes by the induction hypothesis. ■

Theorem 4: Given a maximal bi-Hamiltonian structure associated with the primary vector field v^a, there will exist N scalar functions $H_{(\alpha)}$, that are unique up to constants, pairs of which act as the two Hamiltonians (associated with the two different symplectic forms Ω_{ab} and $\widetilde{\Omega}_{ab}$) for each of the N vector fields $f^a_{(\alpha)}$; more specifically, we have

$$f^a_{(\alpha)} = \Omega^{ab}\,\partial_b H_{(\alpha)} = \widetilde{\Omega}^{ab}\,\partial_b \widetilde{H}_{(\alpha)}, \quad \alpha = 1,2,...,N,$$

with $\widetilde{H}_{(\alpha)} = H_{(\alpha+1)}$ and $H_{(N+1)}$ a linear combination of the $H_{(\alpha)}$.

Proof: First we show that $\mathscr{L}_{S^n v} S = 0$, for all n. We use induction on n, with the case $n = 0$ being true from (3.7) and the definition of S. Let X be an arbitrary vector field. From the Nijenhuis condition we get

$$[S^{n+1}v, SX] = S[S^{n+1}v, X] + S[S^n v, SX] - S^2[S^n v, X].$$

By the induction hypothesis we have $[S^n v, SX] - S[S^n v, X] = 0$. This implies $[S^{n+1}v, SX] = S[S^{n+1}v, X]$, for all vector fields X, which is equivalent to $\mathscr{L}_{S^{n+1}v} S = 0$.

Next we show that this implies $\mathscr{L}_{S^n v}\Omega = 0 = \mathscr{L}_{S^n v}\widetilde{\Omega}$, for all n. From the definition of S we have

$$\widetilde{\Omega}_{ab} = \Omega_{cb} S^c{}_d, \tag{B2}$$

from which we get $(\mathscr{L}_{S^n v}\Omega)_{ab} = (\mathscr{L}_{S^n v}\widetilde{\Omega})_{cb} S^c{}_a$. On the other hand, with $(S^n v)^a = v^a_{(n)} = S^a{}_b v^b_{(n-1)}$ we obtain

$$\begin{aligned}(\mathscr{L}_{S^n v}\Omega)_{ab} &= \Omega_{ab,c} v^c_{(n)} + \Omega_{cb} v^c_{(n),a} + \Omega_{ac} v^c_{(n),b} \\ &= \tfrac{1}{2}\Omega_{[ab,c]} v^c_{(n)} + (\Omega_{cb} v^c_{(n)})_{,a} - (\Omega_{ca} v^c_{(n)})_{,b} \\ &= (\widetilde{\Omega}_{cb} v^c_{(n-1)})_{,a} - (\widetilde{\Omega}_{ca} v^c_{(n-1)})_{,b},\end{aligned}$$

with $\Omega_{[ab,c]} = 0$ and (B2). Since $\widetilde{\Omega}_{[ab,c]} = 0$, we finally obtain

$$\mathscr{L}_{S^n v}\widetilde{\Omega} = \mathscr{L}_{S^{n-1}v}\widetilde{\Omega}. \tag{B3}$$

Taking these relationships together, $\mathscr{L}_{S^{n-1}v}\widetilde{\Omega} = 0$ implies $\mathscr{L}_{S^n v}\widetilde{\Omega} = 0$, and hence $\mathscr{L}_{S^n v}\Omega = 0$. Induction on n establishes the result. The fact that all the iterates of v Lie derive *Om* and *Omt* implies the existence of functions $H_n, \widetilde{H}_n$—unique up to constants—such that $(S^n v)^a = \Omega^{ab}\partial_b H_n = \widetilde{\Omega}^{ab}\partial_b \widetilde{H}_n$. Equation (B3) then tells us that $dH_n = d\widetilde{H}_{n-1}$, and we can choose $H_n = \widetilde{H}_{n-1}$. ■

[1]F. Magri, J. Math. Phys. **19**, 1156 (1979).
[2]A. S. Fokas and B. Fuchssteiner, Lett. Nuovo Cimento **28**, 299 (1980).
[3]A. S. Fokas and B. Fuchssteiner, Physica D **4**, 47 (1981).
[4]I. M. Gel'fand and I. Y. Dorfman, Functional Anal. Appl. **13**, 13 (1979).
[5]I. M. Gel'fand and I. Y. Dorfman, Functional Anal. Appl. **14**, 71 (1980).
[6]A. S. Fokas and P. M. Santini, "A unified approach to recursion operators," preprint, 1988.
[7]V. I. Arnol'd, *Mathematical Methods in Classical Mechanics* (Springer, New York, 1978).
[8]S. Okubo and A. Das, Phys. Lett. B **209**, 311 (1988).
[9]S. Okubo and A. Das, "A systematic study of the Toda lattice," preprint, 1988.
[10]S. Okubo, J. Math. Phys. **30**, 834 (1989).
[11]S. Okubo, J. Math. Phys. **30**, 1043 (1989).
[12]H. Goldstein, *Classical Mechanics*, 2nd ed. (Addison-Wesley, Reading, MA, 1980).
[13]S. Lang, *Algebra* (Addison-Wesley, Reading, MA, 1965).

On the convergence of the Magnus expansion in the Schrödinger representation

Francisco M. Fernández, Julián Echave, and Eduardo A. Castro
Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas (INIFTA), División Química Teórica, Sucursal 4, Casilla de Correo 16, 1900 La Plata, Argentina

(Received 27 April 1989; accepted for publication 6 September 1989)

The convergence properties of the Magnus expansion in the Schrödinger representation are investigated. A quite general result is rigorously derived from first-order perturbation theory. A finite matrix representation is presented for obtaining the exponential time-evolution operator more easily. Two time-dependent models, an oscillator and a spin system, are considered as illustrative examples.

I. INTRODUCTION

There has recently been great interest in the convergence properties of the Magnus expansion[1–3] because it is the basis of the average Hamiltonian theory that proves to be useful in analyzing magnetic resonance experiments[4–8] (and references therein). Since addressing this question in a general rigorous way seems to be a very hard task, some authors have resorted to the investigation of simple models, where large-order calculations are possible.[6–8] For instance, Salzman[7,8] considered problems of the form $H = H_0 + \beta V(t)$, where H_0 and $V(t)$ describe the isolated time-independent system and the time-dependent perturbation, respectively. After studying a number of simple cases, Salzman[7,8] concluded that the Magnus expansion in the Schrödinger representation diverged for all $t > T_0 = 2\pi/\omega_0$, where ω_0 is the natural frequency of the isolated system. When $t = T = 2\pi/\omega$, where ω is the driving frequency, the condition above becomes $\omega > \omega_0$, which severely limits the application of the exponential perturbation theory in the Schrödinger representation.[8]

Salzman's conclusions appear to be β independent.[7,8] On the other hand, the convergence properties of the Magnus expansion are known to depend strongly on the representation.[4,5]

On using an appropriate expansion procedure,[9] Salzman[7,8] obtained a large number of terms of the Magnus expansion for a given order in β and managed to work out the sum exactly. He argued that if the sum of the terms proportional to β diverged, then the series was not useful at higher orders of β.

Some of Salzman's conclusions have recently been confirmed rigorously in the case of the linearly driven harmonic oscillator.[10]

The purpose of the present paper is twofold. First, previous results based on perturbation theory[5,7,8] are more rigorously derived in Sec. II. Second, some properties of Lie algebras are shown in Sec. III to be suitable for investigating the convergence properties of the Magnus expansion. Two simple models are discussed in Sec. IV as illustrative examples.

II. THE MAGNUS EXPANSION

The Magnus expansion has been obtained in more than one way before.[1–3,9] However, for the sake of completeness and in order to introduce useful notation an alternative derivation is presented in this paper.

The Schrödinger equation is a particular case of

$$\frac{d}{dt} U = F\,U, \tag{1}$$

where U and F are linear operators. Without loss of generality the initial condition is chosen to be $U(0) = I$, where I is the identity operator (throughout this paper I will also represent the identity matrix).

The operator $\tilde{y} = U^{-1} y U$ satisfies the following equation of motion:

$$\frac{d}{dt}\tilde{y} = U^{-1}\left(SyF + \frac{d}{dt}y\right) U, \tag{2}$$

where Sy is the superoperator defined over the space of linear operators as $Sy\,x = yx - xy$.

When U is chosen to be of the form

$$U = e^{-A}, \quad A(0) = 0, \tag{3}$$

then

$$\tilde{y} = e^{A} y e^{-A} = y + SAy + \tfrac{1}{2}(SA)^2 y + \cdots = e^{SA} y$$

and Eq. (2) becomes

$$\frac{d}{dt}\tilde{y} = e^{SA}\left(\frac{d}{dt}y + Sy\,F\right) = e^{SA}\left(\frac{d}{dt}y - SF\,y\right), \tag{4}$$

which enables one to define the time derivative of e^{SA} as follows:

$$\frac{d}{dt} e^{SA} = -e^{SA} SF. \tag{5}$$

When $y = A$ ($\tilde{A} = A$), Eq. (4) becomes

$$\frac{d}{dt}\tilde{A} = \frac{d}{dt}A = U^{-1}\left(\frac{d}{dt}A + SA\,F\right) U = e^{SA}\left(\frac{d}{dt}A + SA\,F\right), \tag{6a}$$

which can be solved for $(d/dt)A$ leading to

$$\frac{d}{dt}A = G(SA)F, \tag{6b}$$

where $G(x) = x/(e^{-x} - 1)$. This function can be expanded in Taylor series around the origin: $G(x) = G_0 + G_1 x + \cdots$, where $G_0 = -1$ and

$$G_j = (j+1)^{-1}\left(G_{j-1} + \sum_{k=1}^{j-1} G_k G_{j-k}\right), \quad j>0. \quad (7)$$

The fact enables one to solve Eq. (6b) iteratively and, as a result, the operator A can be expressed as the infinite series

$$A = A_1 + A_2 + \cdots, \quad (8)$$

where A_j is of order j in F.[1-3] The first two terms are

$$A_1 = -\int_0^t F(t')dt',$$
$$A_2 = -\frac{1}{2}\int_0^t \int_0^{t'} [F(t'),F(t'')]dt'\,dt''. \quad (9)$$

The Magnus expansion converges in a neighborhood of $t=0$, where A exists and is differentiable. The convergence interval of the Magnus expansion can, in principle, be obtained as follows. Let $\{\alpha_j, |\alpha_j\rangle\}$ be the set of eigenvalues and eigenfunctions, respectively, of A. For the sake of simplicity we assume that F is anti-Hermitian. Since

$$\langle \alpha_j|(SA)^n F|\alpha_k\rangle = (\alpha_j - \alpha_k)^n \langle \alpha_j|F|\alpha_k\rangle,$$

it is concluded that

$$\left\langle \alpha_j \left| \frac{dA}{dt} \right| \alpha_k \right\rangle = G(\alpha_j - \alpha_k)\langle \alpha_j|F|\alpha_k\rangle, \quad (10)$$

from which it follows that A is not differentiable when $\alpha_j - \alpha_k$ is a multiple of $2\pi i$ and $\langle \alpha_j|F|\alpha_k\rangle \neq 0$. Clearly, the Magnus expansion diverges for all $t > t_c$, where t_c is the smallest time value leading to the above-mentioned condition. Unfortunately, this criterion does not enable one to determine t_c beforehand because the spectrum of the unknown operator A is required.

A reasonable way of estimating t_c has been put forward by Maricq.[5] Assume that F is of the form

$$F = F_0 + \beta F_1(t), \quad (11)$$

where the time-independent operator F_0 has eigenvalues e_j and eigenfunctions $|j\rangle$. For small enough β values, $|\alpha_j\rangle \simeq |j\rangle$ and $\alpha_j \simeq -te_j$. Therefore,

$$t_c \simeq 2\pi/\Delta e, \quad (12)$$

where Δe is the largest value of $|e_j - e_k|$ for which $\langle j|F_1|k\rangle \neq 0$. (This last condition is not always taken into account.[5])

The result above can be obtained in a more rigorous way by means of perturbation theory. We first notice that Eq. (6a) can be rewritten $SU\,dA/dt = (SA\,F)U$. Upon expanding every operator in β-power series (e.g., $A = A^{(0)} + A^{(1)}\beta + \cdots$) we have

$$SU^{(0)}\frac{dA^{(1)}}{dt} - SU^{(1)}F_0$$
$$= SU^{(0)}\frac{d}{dt}A^{(1)} + SF_0U^{(1)}$$
$$= (SA^{(1)}F_0)U^{(0)} + (SA^{(0)}F_1)U^{(0)}$$
$$= -\{SF_0(A^{(1)} + tF_1)\}U^{(0)},$$

because $A^{(0)} = -tF_0$. Therefore

$$\left\langle j\left|\frac{d}{dt}A^{(1)}\right|k\right\rangle = (e_j - e_k)(e^{te_k} - e^{-te_j})^{-1}\langle j|\{U^{(1)} + e^{te_k}(A^{(1)} + tF_1)\}|k\rangle, \quad (13)$$

since $\langle j|SF_0W|k\rangle = (e_j - e_k)\langle j|W|k\rangle$, for every linear operator W. For this reason $A^{(1)}$ is not differentiable when $(e_j - e_k)t$ is a multiple of $2\pi i$ and

$$\langle j|F_1 + t^{-1}A^{(1)} - (te^{te_k})^{-1}U^{(1)}|k\rangle \neq 0.$$

Equation (13) clearly accounts for all the particular results obtained by Salzman[7,8] from simple models. Since the Magnus expansion converges when $\beta = 0$ we can reasonably suppose that $t_c \leqslant 2\pi/\Delta e$.

III. FINITE MATRIX REPRESENTATION

Although the properties of the Lie algebras are known to be useful in studying the dynamics of quantum-mechanical systems,[11,12] some of them are sometimes overlooked. A few such properties relevant to the subject of the present paper are briefly discussed below.

Let L be the Lie algebra spanned by the set $X = \{x_1, x_2, \ldots, x_n\}$ of linearly independent and time-independent operators x_j. All the operators Sx_jx_k belong to L.[11,12] Given an operator B belonging to L we define an n-dimensional square matrix $B' = (B'_{jk})$ as follows:

$$SB\,x_j = \sum_{k=1}^n B'_{jk}x_k, \quad j = 1,2,\ldots,n. \quad (14)$$

If the function $K(z)$ is defined on the spectrum of B', then

$$K(SB)x_j = \sum_{k=1}^n \{K(B')\}_{jk}x_k. \quad (15)$$

Also, $K(B')$ is found to be a polynomial function of B' of degree $s \leqslant n$.[13] When B' has n distinct characteristic values $b_1, b_2, \ldots, b_n$, such a polynomial is of the form[13]

$$K(B') = \sum_{j=1}^n K(b_j) \prod_{\substack{k=1 \\ k\neq j}}^n (b_j - b_k)^{-1}(B' - b_kI). \quad (16)$$

If

$$F = \sum_{j=1}^n f_j(t)x_j, \quad (17)$$

where the $f_j(t)$, $j = 1,2,\ldots,n$, are continuous functions of time, then[11,12]

$$\tilde{x}_j = \sum_{k=1}^n Q'_{jk}(t)x_k, \quad j = 1,2,\ldots,n, \quad (18)$$

where the functions $Q'_{jk}(t)$ satisfy the equations of motion

$$\frac{d}{dt}Q' = -F'Q', \quad Q' = (Q'_{jk}), \quad (19)$$

with the boundary condition $Q'(0) = I$. This last equation is a finite matrix representation of (1), which enables one to handle the problem much more easily.

When F belongs to L the operator A also belongs to L and can be written

$$A = \sum_{j=1}^n a_j(t)x_j, \quad (20)$$

where $a_j(0) = 0$ ($j = 1,2,\ldots,n$). A straightforward calculation using Eqs. (6b) and (15) shows that

$$Q' = e^{A'}, \quad (21)$$

and

$$\frac{d}{dt}a_j = \sum_{k=1}^{n} \{G(A')\}_{kj} f_k, \quad j = 1,2,...,n. \tag{22}$$

It is worth noticing that $G(A') = A'Q'(1-Q')^{-1}$.

This finite matrix representation is useful in investigating the convergence properties of the Magnus expansion for some simple systems. In order to show it, let us look for an operator x belonging to L so that $SA\,x = \lambda x$. If such an operator is found, then λ will be the separation between two eigenvalues of A because

$$\langle\alpha_j|SA\,x|\alpha_k\rangle = (\alpha_j - \alpha_k)\langle\alpha_j|x|\alpha_k\rangle = \lambda\,\langle\alpha_j|x|\alpha_k\rangle.$$

It is not difficult to verify that the separation constant λ and the coefficients of the expansion $x = c_1x_1 + c_2x_2 + \cdots + c_nx_n$ are solutions of the secular equation

$$\sum_{k=1}^{n} (A'_{kj} - \lambda\delta_{kj})c_k = 0, \quad j = 1,2,...,n. \tag{23}$$

In other words, the separation constant is an eigenvalue of A' and the operator A will not be differentiable when λ is a multiple of $2\pi i$ as shown by Eq. (22). ($\lambda = 0$ is always an eigenvalue of A' because $SAx = 0$ when x is proportional to A.)

The dimension of the representation can be reduced when there is a set of operators $Y = \{y_1,y_2,...,y_m\}$, $m < n$, so that Sx_jy_k ($j = 1,2,...,n$, $k = 1,2,...,m$) belong to C_Y (the set of all linear combinations of operators in Y). In such a case, given an operator B belonging to L we may define an m-dimensional square matrix $B'' = (B''_{jk})$ as follows:

$$SB\,y_j = \sum_{k=1}^{m} B''_{jk}y_k, \quad j = 1,2,...,m. \tag{24}$$

Besides,

$$\tilde{y}_j = \sum_{k=1}^{m} Q''_{jk}y_k, \quad j = 1,2,...,m, \tag{25}$$

where

$$\frac{d}{dt}Q'' = -F''Q'', \quad Q'' = (Q''_{jk}), \tag{26}$$

and $Q''(0) = I$. It follows immediately that

$$Q'' = e^{A''}. \tag{27}$$

In the examples of the next section we choose $x_1,x_2,...,x_n$ to be anti-Hermitian operators so that F is anti-Hermitian when the $f_j(t)$ are real functions of time. The Hamiltonian operator $H = iF$ is thus Hermitian and U and A are unitary and anti-Hermitian operators, respectively, when $a_j(t)$ are real functions of time. (The eigenvalues of A are imaginary numbers.)

In order to determine the convergence interval of the Magnus expansion we simply solve the linear equations of motion (26) for increasing time values until an eigenvalue of Q'' equals unity (i.e., an eigenvalue of A'' equals $2\pi i$). When the dimension of the representation cannot be reduced, we use Eq. (19).

The equations developed in this section can be viewed as a generalization of the method proposed by Fernández[10,14] to obtain the exponential form of the time-evolution operator that proved to be useful in studying the convergence of the Magnus expansion for driven oscillators.[10,15] In addition to this, expressions of the form $\exp(uSx_j)x_k$, which are necessary to build the time-evolution operator for collisional problems,[16] are easily derived from the matrix elements of $\exp(x'_j)$.

IV. EXAMPLES

Our first example is the time-dependent oscillator

$$H = \tfrac{1}{2}f_1(t)(p^2 + q^2) + f_2(t)q + f_3(t)p, \tag{28}$$

where $p = -i\,d/dq$. This operator reduces to the one discussed by Salzman[7] and Fernández[10] when $f_1 = \omega_0$, $f_2 = \cos\omega t$, and $f_3 = 0$. Although the Magnus expansion for this problem converges for all t values in the interaction representation,[3] the convergence interval may be finite in the Schrödinger representation as shown below.

The algebra L is spanned by the operators

$$x_1 = -(i/2)(p^2 + q^2), \quad x_2 = -iq, \quad x_3 = -ip, \quad x_4 = -iI. \tag{29}$$

A straightforward calculation using the commutation relations of the operators x_j shows that the nonvanishing matrix elements of A' are

$$A'_{12} = A'_{24} = -a_3, \quad A'_{13} = A'_{34} = a_2, \quad A'_{32} = -A'_{23} = a_1. \tag{30}$$

The eigenvalues of A' are found to be 0 (twofold degenerate) and $\pm ia_1$. Besides, since $A'_{j1} = 0, j = 1,...,4$, it follows that $(A'^n)_{j1} = 0$ and $\{G(A')\}_{j1} = -\delta_{j1}$. For this reason, Eq. (22) leads to

$$a_1(t) = -\int_0^t f_1(t')dt', \tag{31}$$

which shows that the Magnus expansion diverges for all $t \geqslant t_c$, where

$$|a_1(t_c)| = 2\pi. \tag{32}$$

When $f_1 = \omega_0$, it is concluded that $t_c = 2\pi/\omega_0$, disregarding the form of f_2 and f_3. We have thus generalized the results obtained by Salzman[7] and Fernández[10] and shown the usefulness of the finite matrix representation developed in the previous section.

This example is useful to point out a very important fact. The eigenvalues of the operator A are easily shown to be given by

$$\alpha_j = -i\{a_1(j+\tfrac{1}{2}) + a_4 + (a_2^2 + a_3^2)/2a_1\}, \quad j = 0,1,... .$$

Therefore, since $|\alpha_j - \alpha_k|$ can be chosen as large as desired when $t > 0$, the Magnus expansion appears to have a zero convergence radius. However, $\langle\alpha_j|F|\alpha_k\rangle = 0$ if $|j-k| > 1$, and t_c is given by Eq. (32). It is surprising that the condition about the matrix elements of F is frequently omitted.[4,5,8]

The remaining coefficients $a_j(t)$ are more easily obtained by solving the Schrödinger equation in the reduced matrix representation provided by the algebra spanned by $Y = \{x_2,x_3,x_4\}$. Notice that the coefficient $a_4(t)$ cannot be obtained from the finite matrix representations because x_4 commutes with all the other operators. This fact is not a serious drawback because such a coefficient only contributes a trivial phase factor to the time-evolution operator. The calculation of $a_2(t)$ and $a_3(t)$ is not relevant to the present paper.

We next consider the case

$$x_1 = -\frac{i}{2}\sigma_x, \quad x_2 = -\frac{i}{2}\sigma_y, \quad x_3 = -\frac{i}{2}\sigma_z, \tag{33}$$

where σ_x, σ_y, and σ_z are the Pauli spin matrices ($\sigma_j^2 = I$). Since $Sx_1x_2 = -ix_3$, $Sx_2x_3 = -ix_1$, and $Sx_3x_1 = -ix_2$, it is found that the nonvanishing elements of A' are

$$A'^{*}_{12} = A'_{21} = ia_3, \quad A'_{13} = A'^{*}_{31} = ia_2, \quad A'^{*}_{23} = A'_{32} = ia_1, \tag{34}$$

where * stands for complex conjugation. The eigenvalues of A' are 0 and $\pm ia$, where $a = (a_1^2 + a_2^2 + a_3^2)^{1/2}$.

The matrix A' is anti-Hermitian, therefore $G(A')^+ = G(A'^+) = G(-A')$, where $+$ stands for adjoint. Since the coefficients a_j and f_j are real and $G(-SA)A = -A$, it follows from Eq. (22) that

$$\sum_j a_j \frac{d}{dt} a_j = \sum_j \sum_k f_k G(A')_{kj} a_j$$
$$= \sum_j \sum_k a_j G(-A')_{jk} f_k$$
$$= -\sum_k f_k a_k, \tag{35}$$

which finally leads to

$$a(t) \leqslant \int_0^t f(t')dt', \tag{36}$$

where $f = (f_1^2 + f_2^2 + f_3^2)^{1/2}$, in agreement with Maricq.[17]

V. COMMENTS

The main result of Sec. II [i.e., Eq. (13)] shows that the conclusion obtained by Salzman[7,8] from particular examples is quite general and that the applicability of the methods based on the Magnus expansion is rather limited.

The finite matrix representation developed in Sec. III greatly facilitates the calculation of the exponential time-evolution operator and thereby the investigation of the convergence properties of the Magnus expansion. More complex problems than those in Sec. IV will be treated elsewhere in a forthcoming paper.

The accurate determination of the convergence interval of the Magnus expansion is of utmost importance in most physical applications because it would help to choose a more appropriate representation of the Schrödinger equation.[5]

1W. Magnus, Commun. Pure Appl. Math. 7, 649 (1954).
2D. W. Robinson, Helv. Phys. Acta 36, 140 (1963).
3P. Pechukas and J. C. Light, J. Chem. Phys. 44, 3897 (1966).
4M. M. Maricq, Phys. Rev. B 25, 6622 (1982).
5M. M. Maricq, J. Chem. Phys. 85, 5167 (1986).
6E. B. Fel'dman, Phys. Lett. A 104, 479 (1984).
7W. R. Salzman, Chem. Phys. Lett. 124, 531 (1986); J. Chem. Phys. 85, 4605 (1986).
8W. R. Salzman, Phys. Rev. A 36, 5074.(1987).
9W. R. Salzman, J. Chem. Phys. 82, 822 (1985).
10F. M. Fernández, J. Chem. Phys. 88, 490 (1988).
11J. Wei and E. Norman, J. Math. Phys. 4, 575 (1963).
12R. Gilmore, *Lie Groups, Lie Algebras, and Some of their Applications* (Wiley–Interscience, New York, 1974).
13F. R. Gantmacher, *The Theory of Matrices* (Chelsea, New York, 1959), Vol. I, Chap. V.
14F. M. Fernández, J. Math. Phys. 28, 2908 (1987).
15F. M. Fernández, J. Echave, and E. A. Castro, Chem. Phys. 117, 101 (1987).
16B. Gazdy and D. A. Micha, J. Chem. Phys. 82, 4926 (1985).
17M. M. Maricq, J. Chem. Phys. 86, 5647 (1987).

On the convergence of the Feynman path integral for a certain class of potentials

Vassilis G. Papanicolaou
Department of Mathematics, Duke University, Durham, North Carolina 27706

(Received 21 March 1989; accepted for publication 4 October 1989)

A direct proof is given of the convergence of the discretized Feynman path integral to the fundamental solution of the (time-dependent) Schrödinger equation, for potentials which are bounded, integrable and continuous on the real line. No further smoothness assumptions are required.

I. INTRODUCTION. THE SCHRÖDINGER EQUATION AND THE DISCRETIZED FEYNMAN PATH INTEGRAL

The one-dimensional Schrödinger equation, after a real scaling of x and t, can be written in the reduced (dimensionless) form

$$\frac{1}{i}\frac{\partial k}{\partial t}=\frac{1}{2}\frac{\partial^2 k}{\partial x^2}+q(x)k, \quad t\in(0,\infty), \quad x\in(-\infty,\infty), \tag{1}$$

(where $-q$ is what is usually called "the potential"). We are interested in the fundamental solution (Green's function) $k(t,x,y)$ of (1), namely, the solution with initial condition

$$k(0,x,y)=\delta(x-y), \tag{2}$$

in the sense that, if $w(t,x)$ is a solution of (1) with

$$w(0,x)=f(x), \quad f\in L^2(R^1),$$

then, for all $t>0$,

$$w(t,x)=\int_{-\infty}^{\infty} k(t,x,y)f(y)dy,$$

and if f is smooth enough,

$$f(x)=\lim_{t\downarrow 0}\int_{-\infty}^{\infty} k(t,x,y)f(y)dy.$$

If we set

$$L=\frac{1}{2}\frac{d^2}{dx^2}+q,$$

then the function $k(t,x,y)$ is the kernel of the unitary operator e^{itL}. One consequence of that is that $k(t,x,y)$ must be symmetric in x and y. In the finite interval case $a\leqslant x\leqslant b$, with certain boundary conditions, $k(t,x,y)$ exists only in the distributional sense, since unitary operators are not compact. But in the infinite case (our case) $k(t,x,y)$ can be bounded and continuous. If the operator L is of the limit point type at $\pm\infty$ (which is the case if, for example, q is bounded), then $k(t,x,y)$ is unique.

Feynman conjectured that k is given by

$$k(t,x,y)=\lim_{m\to\infty} K_m(t,x,y), \tag{3a}$$

where K_m is the discretized path integral

$$K_m(t,x,y) =\left(\frac{m}{2\pi i t}\right)^{m/2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left[i\sum_{j=1}^{m}\frac{(z_j-z_{j-1})^2}{2\Delta t}\right] \times\exp\left[i\sum_{j=1}^{m-1}q(z_j)\Delta t\right]dz_1\cdots dz_{m-1}, \tag{3b}$$

with $z_0=x$, $z_m=y$, $\Delta t=t/m$, $m\geqslant 2$. The expression $i^{1/2}$ is meant to be $e^{i\pi/4}$ and the integrals are in the improper (Riemann) sense, namely,

$$\int_{-\infty}^{\infty} f(x)dx=\lim_{M,N\to\infty}\int_{-M}^{N} f(x)dx. \tag{3c}$$

For the cases $q(x)=c$, cx, or cx^2 (c is a constant), formula (3a) can be verified directly. The case $q(x)=cx^2$ is the most interesting.[1] Fujiwara[2,3] showed the validity of (3a) for a class of C^∞ potentials. Nelson[4] showed that, under very general assumptions for q, the operators with kernels $K_m(t,x,y)$ converge, as $m\to\infty$, to the operator e^{itL}, in the strong operator topology, but this is not enough to prove (3a).

In this work we give the proof of (3a) for $q\in LC_b(R^1)$, i.e., for q bounded, continuous, and integrable on the line (not necessarily real valued). As far as we know, all existing proofs require q to be C^∞ and they depend on Fourier transform techniques (see also Ref. 1). Our way is direct and we do not need more smoothness. Notice that $LC_b(R^1)$ is a Banach space with norm $\|\cdot\|=\max\{\|\cdot\|_1,\|\cdot\|_\infty\}$ (say). Moreover, if $q\in LC_b(R^1)$ and $p\geqslant 1$, then q^p and e^q-1 are both in $LC_b(R^1)$. In fact

$$\|q^p\|_1\leqslant\|q\|_\infty^{p-1}\|q\|_1 \tag{4a}$$

and

$$\|e^q-1\|_1\leqslant\left[\frac{(e^{\|q\|_\infty}-1)}{\|q\|_\infty}\right]\|q\|_1\leqslant e^{\|q\|_\infty}\|q\|_1. \tag{4b}$$

II. THE BORN EXPANSION

There is a way to construct the fundamental solution of (1) as an infinite series. This series is usually called Born expansion and it is a perturbation series near $q\equiv 0$.

We start with the fundamental solution for the case $q\equiv 0$, namely,

$$k_0(t,x,y)=\frac{1}{\sqrt{2\pi i t}}\exp\left[\frac{i(x-y)^2}{2t}\right] \quad (\sqrt{i}=e^{i\pi/4}), \tag{5a}$$

 0022-2488/90/020342-06$03.00

and, for $n = 1,2,...$, we set

$$k_n(t,x,y) = i\int_0^t\int_{-\infty}^{\infty} k_0(s,x,z)q(z) \times k_{n-1}(t-s,z,y)dz\,ds, \quad (5b)$$

or equivalently

$$k_n(t,x,y) = \int_0^t\int_0^{t-s_1}\cdots\int_0^{t-s_1-\cdots-s_{n-1}} H_n[s_1,\cdots,s_n](t,x,y) \times ds_n\cdots ds_2\,ds_1, \quad (5c)$$

where

$$H_n[s_1,\cdots,s_n](t,x,y) = i^n\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} k_0(s_1,x,z_1)\cdots k_0(s_n,z_{n-1},z_n) \times k_0(t-s_1-\cdots-s_n,z_n,y)q(z_1)\cdots q(z_n)dz_n\cdots dz_1. \quad (5d)$$

Then

$$k(t,x,y) = \sum_{n=0}^{\infty} k_n(t,x,y). \quad (6)$$

[This can be viewed as an analytic continuation of the fundamental solution of the parabolic equation obtained from (1) by dropping the factor $1/i$].

Each $k_n(t,x,y)$ is symmetric in x and y by induction on (5b). The assumption $q\in LC_b(R^1)$ together with dominated convergence imply that $k_n(t,x,y)$ is continuous in (t,x,y). Furthermore, by (5b), for $n\geqslant 1$ we have

$$|k_n(t,x,y)|\leqslant\|q\|_1^n(1/\sqrt{2\pi})^{n+1}M_n(t), \quad (7)$$

where

$$M_0(t) = 1/\sqrt{t} \quad (8a)$$

and, for $n\geqslant 1$,

$$M_n(t) = \int_0^t \frac{M_{n-1}(t-s)}{\sqrt{s}}ds. \quad (8b)$$

In other words [as in (5c) and (5d)],

$$M_n(t) = \int_0^t\int_0^{t-s_1}\cdots\int_0^{t-s_1-\cdots-s_{n-1}} \frac{ds_n\cdots ds_2\,ds_1}{\sqrt{s_1s_2\cdots s_n(t-s_1-\cdots-s_n)}}. \quad (8c)$$

Lemma 1: Formulas (8a) and (8b) imply

$$M_n(t) = A_n t^{(n-1)/2}, \quad (9a)$$

where

$$A_{2r} = (4\pi)^r r!/(2r)! \text{ and } A_{2r+1} = \pi^{r+1}/r!. \quad (9b)$$

In particular

$$\lim_n(A_{n+1}/A_n) = 0. \quad (9c)$$

Proof: Elementary calculus exercise. ■

Remark: By using Stirling's formula in (9b) we obtain

$A_{n+1}/A_n = O(1/\sqrt{n})$.

An immediate corollary is that the series in (6) converges absolutely and uniformly in (t,x,y) provided that $0<\epsilon\leqslant t\leqslant b<\infty$. Therefore, $k(t,x,y)$ of (6) is continuous on $(0,\infty)\times R^1\times R^1$ and symmetric in x and y. In fact, (7) implies that $k(t,x,y)$ is uniformly bounded if we restrict t in some interval of the form $[\epsilon,b]$. To see that it is the fundamental solution, of (1), at least in the weak sense, we notice the following.

For $n\geqslant 1$, Eq. (5b) implies (we have used the notation D_1 and D_{22} in a few places below to denote partial derivatives, because our usual notation could cause confusion)

$$\frac{1}{i}\frac{\partial k_n(t,x,y)}{\partial t} = q(x)k_{n-1}(t,x,y) + \int_0^t\int_{-\infty}^{\infty}[D_1k_0(s,x,z) \times q(z)k_{n-1}(t-s,z,y)]dz\,ds,$$

and

$$\frac{1}{2}\frac{\partial^2 k_n(t,x,y)}{\partial x^2} = i\int_0^t\int_{-\infty}^{\infty}\frac{1}{2}D_{22}k_0(s,x,z)q(z) \times k_{n-1}(t-s,z,y)dz\,ds = \int_0^t\int_{-\infty}^{\infty}[D_1k_0(s,x,z)q(z) \times k_{n-1}(t-s,z,y)]dz\,ds,$$

since, by (5a),

$$\frac{1}{i}\frac{\partial k_0(t,x,y)}{\partial t} = \frac{1}{2}\frac{\partial^2 k_0(t,x,y)}{\partial x^2}.$$

By summing the first two equations for all $n\geqslant 1$, and then by comparing their right-hand sides (using at the same time the third equation), we conclude that $k(t,x,y)$ satisfies (1). To assure convergence we need $x^2q(x)$ to be integrable on the line and in this case, $k(t,x,y)$ is a strong solution of (1). Since such q's are dense in $LC_b(R^1)$, we can establish the general case by dominated convergence, but, in general, our $k(t,x,y)$ is only a weak solution of (1). To verify the initial condition (2) we just observe that

$$k_0(0,x,y) = \delta(x-y),$$

while, by (5b),

$$k_n(0,x,y) = 0 \quad n\geqslant 1.$$

Remarks: It is a standard calculus exercise that

$$\int_{-\infty}^{\infty} k_0(t,x,y)dy = 1, \quad (10)$$

which implies immediately that $k_0(t,x,y)$ satisfies the Chapman–Kolmogorov (matrix multiplication) equation

$$k_0(s+t,x,y) = \int_{-\infty}^{\infty} k_0(s,x,z)k_0(t,z,y)dz. \quad (11)$$

Both (10) and (11) are in the sense of (3c). Equation (11) is crucial for what follows. Since $k(t,x,y)$ is the kernel of e^{itL}, namely the analytic continuation (in t) of a heat kernel, it must also satisfy

$$k(s+t,x,y) = \int_{-\infty}^{\infty} k(s,x,z)k(t,z,y)dz, \quad (12)$$

in a certain improper sense.

Can we extend the analysis of this section to the multidimensional case $x \in R^d, d \geqslant 2$? In this case $|k_0(t,x,y)| = (2\pi t)^{-d/2}$, therefore we cannot use the same techniques [namely (7) and Lemma 1] to get bounds for $k_n(t,x,y)$ that guarantee the convergence of the series in (6). But, possibly, these bounds can be achieved by allowing q to be smoother (see also Ref. 3).

III. THE CONVERGENCE OF THE DISCRETIZED PATH INTEGRAL

We start with a small lemma which justifies the interchange of summation and integration for certain improper integrals.

Lemma 2: Let $f_j \in LC_b(R^1)$ and $\Delta_j > 0, j = 1,\cdots,m-1$. Then

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[i \sum_{j=1}^{m} \frac{(z_j - z_{j-1})^2}{2\Delta_j}\right] \times \exp\left[\sum_{j=1}^{m-1} f_j(z_j)\right] dz_1 \cdots dz_{m-1}$$
$$= \sum_{n=0}^{\infty} \frac{1}{n!} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[i \sum_{j=1}^{m} \frac{(z_j - z_{j-1})^2}{2\Delta_j}\right] \times \left[\sum_{j=1}^{m-1} f_j(z_j)\right]^n dz_1 \cdots dz_{m-1}, \qquad (13)$$

where the multiple integrals are in the (improper) sense of (3c).

Proof: Induction on m. For $m = 2$ we have

$$\exp\left[\sum_{j=1}^{m-1} f_j(z_j)\right] = e^{f_1(z_1)} = 1 + [e^{f_1(z_1)} - 1],$$

but $e^{f_1} - 1$ is in $L^1(R^1)$ by (4b) and hence, one can apply the Dominated Convergence Theorem to get (13). The contribution of the term 1 above is finite because of (10) and (11).

The same idea works for the general case, namely,

$$\exp\left[\sum_{j=1}^{m-1} f_j(z_j)\right] = \prod_{j=1}^{m-1} [e^{f_j(z_j)} - 1] + \Sigma + (-1)^m,$$

where the product is in $L^1(R^{m-1})$ and Σ is a finite sum of terms of the form

$$\exp\left[\sum_{j \in S(m-1)} f_j(z_j)\right],$$

with $S(m-1)$ being a proper subset of $\{1,2,\ldots,m-1\}$. Thus, the inductive hypothesis can be applied to Σ. Notice that we have used (10) and (11) again. ■

Lemma 2 implies that $K_m(t,x,y)$ of (3b) can be written as

$$K_m(t,x,y) = \sum_{n=0}^{\infty} K_{m,n}(t,x,y), \qquad (14a)$$

where

$$K_{m,n}(t,x,y) = \frac{1}{n!}\left(\frac{m}{2\pi i t}\right)^{m/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[i \sum_{j=1}^{m} \frac{(z_j - z_{j-1})^2}{2\Delta t}\right] \times \left[i \sum_{j=1}^{m-1} q(z_j)\Delta t\right]^n dz_1 \cdots dz_{m-1}. \qquad (14b)$$

To finish the proof of (3a), we need to show that, in (14a), it is permissible to bring $\lim_m$ inside the summation, namely,

$$\lim_m K_m(t,x,y) = \sum_{n=0}^{\infty} \lim_m K_{m,n}(t,x,y) \qquad (15)$$

and also that

$$\lim_m K_{m,n}(t,x,y) = k_n(t,x,y), \qquad (16)$$

where $k_n(t,x,y)$ is defined by the formulas (5). In fact, the central idea of this work is that, for any fixed $n \geqslant 0$, $K_{m,n}(t,x,y)$ is approximately a Riemann sum associated to $k_n(t,x,y)$ of (5c).

Proof of (15) and (16): Formula (15) is valid if, for each $t > 0$, we can find numbers $B_n(t)$, $n \geqslant 1$, independent of m, such that

$$|K_{m,n}(t,x,y)| \leqslant B_n(t), \quad \text{for all } m \geqslant 2 \qquad (17)$$

and

$$\sum_{n=0}^{\infty} B_n(t) < \infty, \quad \text{for all } t > 0. \qquad (18)$$

We set

$$\left[\sum_{j=1}^{m-1} q(z_j)\right]^n = E_1 + \cdots + E_n, \qquad (19)$$

where E_k consists of the terms of the expansion with exactly k distinct factors. For example

$$E_1 = q(z_1)^n + \cdots + q(z_{m-1})^n,$$
$$E_2 = q(z_1)^{n-1}q(z_2) + q(z_1)^{n-2}q(z_2)^2 + \cdots + q(z_{m-2})q(z_{m-1})^{n-1}, \text{ etc.}$$

In general we can write

$$E_k = \sum_{1 \leqslant j_1 < \cdots j_k \leqslant m-1} \left[\sum_{\substack{n_1 + \cdots n_k = n \\ n_l \geqslant 1,\, l=1,\ldots,k}} \frac{n!}{n_1! \cdots n_k!} \times q(z_{j_1})^{n_1} \cdots q(z_{j_k})^{n_k}\right]. \qquad (20)$$

Notice that, if $m - 1 < n$, then $E_k = 0$ for $k > m - 1$.

Next we set (k, here, is a superscript)

$$N^k_{m,n}(t,x,y) = \frac{(i\Delta t)^n}{n!}\left(\frac{m}{2\pi i t}\right)^{m/2} \times \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[i \sum_{j=1}^{m} \frac{(z_j - z_{j-1})^2}{2\Delta t}\right] E_k \, dz_1 \cdots dz_{m-1}, \qquad (21)$$

and so

$$K_{m,n}(t,x,y) = \sum_{k=1}^{n} N^k_{m,n}(t,x,y). \qquad (22)$$

Here is our main estimate.

Lemma 3: The quantities $N^k_{m,n}(t,x,y)$, defined by (21) and (20), satisfy

$$|N^k_{m,n}(t,x,y)| \leqslant (n^n/n!) Q^n \Delta t^{n-k} A_k t^{(k-1)/2}, \qquad (23)$$

with $Q = \max(\|q\|_\infty, \|q\|_1)$ and A_k given by (9b).

Proof: Let $m\geqslant 2$, n, and $k\leqslant n$ be fixed positive integers. For $l=1,...,k$ we choose integers $n_l\geqslant 1$ and j_l, such that $n_1+\cdots+n_k=n$ and $1\leqslant j_1<\cdots<j_k\leqslant m-1$. By (5a) and (11) we have

$$\left|\frac{(i\Delta t)^n}{n!}\left(\frac{m}{2\pi i t}\right)^{m/2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left[i\sum_{j=1}^{m}\frac{(z_j-z_{j-1})^2}{2\Delta t}\right]q(z_{j_1})^{n_1}\cdots q(z_{j_k})^{n_k}dz_1\cdots dz_{m-1}\right|$$
$$=\left|\frac{\Delta t^n}{n!}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}J(j_1,\cdots j_k,z_{j1},...,z_{jk},m,\Delta t,x,y)q(z_{j_1})^{n_1}\cdots q(z_{jk})^{n_k}dz_{j_1}\cdots dz_{j_k}\right|, \tag{24a}$$

where

$$J(j_1,...j_k,z_{j_1},...,z_{jk},m,\Delta t,x,y)=k_0(j_1\Delta t,x,z_{j_1})k_0((j_2-j_1)\Delta t,z_{j_1},z_{j_2})\cdots k_0((m-j_k)\Delta t,z_{jk},y). \tag{24b}$$

Thus, involving (4a), the expression in (24a) is bounded by

$$\frac{\Delta t^n}{n!}\frac{1}{\sqrt{2\pi j_1\Delta t}}\frac{1}{\sqrt{2\pi(j_2-j_1)\Delta t}}\cdots\frac{1}{\sqrt{2\pi(m-j_k)\Delta t}}\int_{-\infty}^{\infty}|q(z)|^{n_1}\cdots\int_{-\infty}^{\infty}|q(z)|^{n_k}dz$$
$$\leqslant\frac{\Delta t^n}{n!}\frac{1}{\sqrt{j_1\Delta t}}\frac{1}{\sqrt{(j_2-j_1)\Delta t}}\cdots\frac{1}{\sqrt{(m-j_k)\Delta t}}\|q\|_\infty^{n-k}\|q\|_1^k$$
$$\leqslant\frac{Q^n\Delta t^n}{n!}\frac{1}{\sqrt{j_1\Delta t}}\frac{1}{\sqrt{(j_2-j_1)\Delta t}}\cdots\frac{1}{\sqrt{(m-j_k)\Delta t}}, \tag{24c}$$

where $t>0$ and $Q=\max(\|q\|_\infty,\|q\|_1)$. Observe that the bound in (24c) is independent of the n_l's.

To continue, we need a trivial estimate, namely,

$$\sum_{\substack{n_1+\cdots+n_k=n\\ n_l\geqslant 1,l=1,...,k}}\frac{n!}{n_1!\cdots n_k!}\leqslant n^n. \tag{25}$$

Combining (20), (21), (24), and (25) we obtain

$$|N_{m,n}^k(t,x,y)|\leqslant(n^n/n!)Q^n\Delta t^{n-k}M_m^k(t), \tag{26}$$

where we have set (for convenience)

$$M_m^k(t)=\sum_{1\leqslant j_1<\cdots j_k\leqslant m-1}\frac{1}{\sqrt{j_1\Delta t}}\frac{1}{\sqrt{(j_2-j_1)\Delta t}}\cdots\frac{1}{\sqrt{(m-j_k)\Delta t}}\Delta t^k \tag{27}$$

(again, k is a superscript). But $M_m^k(t)$ is a Riemann sum associated to $M_k(t)$ of (8c). In fact, for all k and all $m\geqslant 2$,

$$M_m^k(t)\leqslant M_k(t)=A_kt^{(k-1)/2}, \tag{28}$$

where A_k is given by (9b) [for the proof of (28) see the Appendix]. We combine the above formula with (26) and the lemma is proved. ■

Corollary 1: There are constants C and B such that

$$|K_{m,n}(t,x,y)|\leqslant CB^n[t^{(n-1)/2}+1/\sqrt{t}](\Delta t^{[n/2]}+A_{[n/2]}), \tag{29}$$

for all $m\geqslant 2,t>0,x,y\in R^1$. Here, $[n/2]$ is the greatest integer $\leqslant n/2$.

In particular,

$$\sum_{n=0}^{\infty}|K_{m,n}(t,x,y)|\leqslant B(t)<\infty, \tag{30}$$

where $B(t)$ depends only on t.

Proof: By (22) and (23)

$$|K_{m,n}(t,x,y)|\leqslant\frac{n^n}{n!}Q^n\left[t^{(n-1)/2}+\frac{1}{\sqrt{t}}\right]\sum_{k=1}^{n}A_k\Delta t^{n-k}.$$

Let $A=\Sigma_{k=0}^{\infty}A_k$. This is finite, by (9b). Thus, the above inequality becomes (the factor $n^n/n!$ is estimated by Stirling's formula; C and B are constants)

$$|K_{m,n}(t,x,y)|\leqslant CB^n[t^{(n-1)/2}+1/\sqrt{t}]$$
$$\times\left(\sum_{k=1}^{[n/2]}A_k\Delta t^{n-k}+\sum_{k=[n/2]+1}^{n}A_k\Delta t^{n-k}\right)$$
$$\leqslant CB^n[t^{(n-1)/2}+1/\sqrt{t}]$$
$$\times(A_n/2)(\Delta t^{[n/2]}+A_{[n/2]}),$$

since A_k of (9c) is eventually decreasing. We can absorb the factor $An/2$ in CB^n by increasing C and B a little, hence (29) is proved.

If $m>B^2t$, then (30) follows from (29) because $\Delta t=t/m$ and $A_{n+1}/A_n\to 0$. For the smaller m's, (30) is true by (14b). ■

Corollary 2: For each $t>0,x,y\in R^1$ and $n\geqslant 0$,

$$\lim_m K_{m,n}(t,x,y)=k_n(t,x,y). \tag{31}$$

Proof: If $k<n$, (23) implies (because of the factor Δt^{n-k} in the right-hand side) that

$$\lim_m N_{m,n}^k(t,x,y)=0.$$

Therefore, by (22),

$$\lim_m K_{m,n}(t,x,y)=\lim_m N_{m,n}^n(t,x,y).$$

But, from (21) and (20) we get (without loss of generality $m-1\geqslant n$, since n is fixed and $m\to\infty$)

$$N^{n}_{m,n}(t,x,y) = i^n \Delta t^n \sum_{1 \leqslant j_1 < \cdots j_n \leqslant m-1} \left(\frac{m}{2\pi i t}\right)^{m/2}$$
$$\times \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[i \sum_{j=1}^{m} \frac{(z_j - z_{j-1})^2}{2\Delta t} \right] q(z_{j_1}) \cdots q(z_{j_n}) dz_1 \cdots dz_{m-1}.$$

Using (5a) and (11) as in the proof of Lemma 3, we obtain [see (24b) for the definition of J below]

$$N^{n}_{m,n}(t,x,y) = i^n \Delta t^n \sum_{1 \leqslant j_1 < \cdots j_n \leqslant m-1} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} J(j_1, \cdots j_n, z_{j_1}, \cdots, z_{j_n}, m, \Delta t, x, y) q(z_{j_1}) \cdots q(z_{j_n}) dz_{j_1} \cdots dz_{j_n},$$

which is a Riemann sum associated to $k_n(t,x,y)$ of (5c) and (5d). ■

Therefore, we have established the following.

Theorem: For $q \in LC_b(R^1)$ let $k(t,x,y)$ be the fundamental solution of the Schrödinger equation (1), as explained in Sec. I. The associated discretized path integral is $(m \geqslant 2)$

$$K_m(t,x,y) = \left(\frac{m}{2\pi i t}\right)^{m/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[i \sum_{j=1}^{m} \frac{(z_j - z_{j-1})^2}{2\Delta t} \right] \times \exp\left[i \sum_{j=1}^{m-1} q(z_j) \Delta t \right] dz_1 \cdots dz_{m-1},$$

with $z_0 = x$, $z_m = y$, $\Delta t = t/m$, $m \geqslant 2$ and $i^{1/2} = e^{i\pi/4}$ [the integrals are in the improper (Riemann) sense]. Then, for each $t > 0$, there is a number $B(t) < \infty$ (depending only on t) such that, for all m, x, and y,

$$|K_m(t,x,y)| \leqslant B(t),$$

and

$$\lim_m K_m(t,x,y) = k(t,x,y)$$

pointwise in (t,x,y).

Remarks: For $q \in LC_b(R^1)$, our result implies the earlier known fact (see for example Sec. 4 of Ref. 4) that the operators with kernels $K_m(t,x,y)$, acting on $L^2(R^1)$, converge to e^{itL} [where $L = \frac{1}{2}(d^2/dx^2) + q$, as in Sec. I] in the strong operator topology.

ACKNOWLEDGMENTS

We want to thank Professor S. Venakides for showing to us the insightful paper of J. B. Keller and D. W. McLaughlin (Ref. 4) which was the initiative of this work. Also, we want to thank Professor G. Kallianpur for giving us a lot of information of what is known for this kind of problems and Professor J. B. Keller for his encouragement and suggestions.

APPENDIX

Here we give the proof of (28), namely that

$$M^k_m(t) \leqslant A_k t^{(k-1)/2}, \quad t > 0,$$

where $M^k_m(t)$ is given by (27) and A_k by (9b). Notice that $M^k_m(t) = 0$, if $k \geqslant m$, because the sum in (27) is empty.

For $k \geqslant 1$ we set

$$F_k(m) = \sum_{1 \leqslant j_1 < \cdots j_k \leqslant m-1} \frac{1}{\sqrt{j_1}} \frac{1}{\sqrt{j_2 - j_1}} \cdots \frac{1}{\sqrt{m - j_k}}. \quad \text{(A1)}$$

Therefore, (27) becomes

$$M^k_m(t) = \Delta t^{(k-1)/2} F_k(m). \quad \text{(A2)}$$

But, for $k \geqslant 2$, by setting $l_r = j_{r+1} - j_1$, $r = 1, \ldots, k-1$, we obtain

$$F_k(m) \leqslant \sum_{1 \leqslant j_1 \leqslant m-1} \frac{1}{\sqrt{j_1}} \left(\sum_{1 \leqslant l_1 < \cdots < l_{k-1} \leqslant m - j_1 - 1} \times \frac{1}{\sqrt{l_1}} \frac{1}{\sqrt{l_2 - l_1}} \cdots \frac{1}{\sqrt{m - j_1 - l_{k-1}}} \right)$$

or

$$F_k(m) \leqslant \sum_{j=1}^{m-1} \frac{F_{k-1}(m-j)}{\sqrt{j}}. \quad \text{(A3)}$$

If we set $F_0(m) = 1/\sqrt{m}$, the above the formula is true even for $k = 1$ (in fact, it becomes an equality).

To prove (28), it's enough to show [because of (A2) and the fact that $\Delta t = t/m$] that

$$F_k(m) \leqslant A_k m^{(k-1)/2}. \quad \text{(A4)}$$

Let us use induction. Formula (A4) becomes an equality if $k = 0$. To finish the proof, because of (A3), it suffices to show that

$$\sum_{j=1}^{m-1} \frac{A_{k-1}(m-j)^{(k-2)/2}}{\sqrt{j}} \leqslant A_k m^{(k-1)/2}$$

or equivalently

$$\sum_{j=1}^{m-1} \frac{(1 - j/m)^{(k-2)/2}}{\sqrt{j/m}} \frac{1}{m} \leqslant \frac{A_k}{A_{k-1}}.$$

Now, the left-hand side above is a lower Riemann sum of

$$I = \int_0^1 \frac{(1-x)^{(k-2)/2}}{\sqrt{x}} dx$$

[since $f_k(x) = (1-x)^{(k-2)/2}/\sqrt{x}$ is decreasing in (0,1), if $k \geqslant 2$, and has only one critical point at $x = 1/2$, if $k = 1$]. By substituting $x = \sin^2 \phi$ and I and using elementary calculus (Wallis formulas) we get

$$I = A_k / A_{k-1},$$

which finishes the argument.

[1] G. Kallianpur, D. Kannan, and R. L. Karandikar, Ann. Inst. Henri Poincarè **21**, 323–361 (1985).
[2] D. Fujiwara, Duke Math. J. **47**, 559–600 (1980).
[3] D. Fujiwara, in *Theory and Applications of Random Fields, Lecture Notes in Control and Information Sciences No. 49*, edited by G. Kallianpur (Springer, New York, 1983), pp. 40–48.
[4] J. B. Keller and D. W. McLaughlin, Am. Math. Monthly **82**, 451–465 (1975).

Untenability of a proposed constrained dynamics for the damped harmonic oscillator

Nivaldo A. Lemos
Instituto de Física, Universidade Federal Fluminense, 24020 Niterói, RJ, Brazil

(Received 3 August 1988; accepted for publication 20 September 1989)

It is argued that the method introduced recently by Wang [J. Phys. A: Math. Gen. **20**, 4745 (1987)] to quantize the damped harmonic oscillator is untenable because it does not reproduce the standard results for the quantum oscillator in the limit of arbitrarily small damping. The method is also shown to be inconsistent if the friction coefficient is allowed to take any positive value.

I. INTRODUCTION

The great importance of constrained Hamiltonian systems is witnessed by their occurrence in modern attempts to describe the fundamental interactions of elementary particles. The canonical formulation of gauge and string theories for instance, leads naturally to constraints and Dirac's formalism[1] to deal with them is usually invoked. Sometimes it is useful to artificially introduce constraints into certain systems with the intention of bringing forth new symmetries that may simplify the analysis of the dynamics of such systems. This is customarily done at the cost of an enlargement of the phase space of the system through additional degrees of freedom that are eventually eliminated. Another way of achieving it is by making use of Lagrangians depending on derivatives of higher than first order and thus working in the realm of the so-called generalized dynamics. In the case of the harmonic oscillator for example, a second-order Lagrangian[2] can be chosen so that it furnishes the correct equation of motion. It has been claimed[2] that the subsequent quantization on the basis of such a Lagrangian yields results that are not equivalent to the standard ones, and this has led to the proposal of unorthodox quantization procedures to cure what was viewed as a disease.[3] A careful treatment of the constraints however, has subsequently shown[4] that those claims were groundless and that there is no need to change the canonical quantization procedure in the framework of generalized dynamics.

The phenomenological quantization of dissipative systems is being investigated from several different points of view. The canonical quantization based on Bateman's time-dependent Lagrangian[5,6] has been criticized as actually referring to a variable-mass system.[7,8] Then followed other suggestions to attack the problem, such as taking the mass as a dynamical variable,[9] or introducing nonlinear Schrödinger equations.[10–15] The inclusion of an external stochastic force has been studied, chiefly to overcome certain difficulties engendered by Bateman's Lagrangian,[16,17] while the exploitation of the quantum Liouville equation has also been tried.[18] Since the variety of approaches is considerably wide, the reader is referred to Dekker's review[19] where an extensive survey of the literature can be found. As to the phenomenological canonical quantization of dissipative (more generally, nonconservative) systems, it has been repeatedly shown to be impossible or ambiguous.[20–25]

Recently, a new treatment[26] of the damped harmonic oscillator has been put forward, regarding it to be a constrained system described by a constrained generalized Hamiltonian. To our knowledge, this is the first attempt so far to construct a constrained Hamiltonian model for the description of a dissipative system at the quantum level. Instead of artificially enlarging the phase space, as has been done for the harmonic oscillator,[2–4] Wang[26] introduces constraints in the ordinary phase space in such a way that the state vector has to satisfy a nonlinear subsidiary condition, in addition to a linear Schrödinger equation. What we undertake to prove in the present paper is that Wang's theory is untenable on physical grounds, particularly because it does not reproduce standard results for the quantum oscillator in the limit of a vanishing friction coefficient. We shall also adduce some general arguments to the effect that any other theory constructed along the lines suggested by Wang is bound to fail.

II. A PROPOSED CONSTRAINED DYNAMICS FOR THE QUANTIZED DAMPED HARMONIC OSCILLATOR

In a recent paper, Wang[26] proposed a new and ingenious method to quantize the damped harmonic oscillator by considering the equation of motion.

$$\ddot{x} + \gamma\dot{x} + \omega^2 x = 0, \tag{1}$$

as arising from a constrained Hamiltonian system in the bidimensional phase space (x,p). His classical treatment introduces a first-class constraint ϕ_1 in phase space. According to Dirac's theory of constrained systems,[1] in the quantized theory the first-class constraints must be imposed as supplementary conditions on the physical state belonging to the Hilbert space $\mathscr{H} = L^2(\mathbb{R})$. Let $\hat{H}_0$ be the Hamiltonian operator of a harmonic oscillator of frequency ω, that is,

$$\hat{H}_0 = \hat{p}^2/2m + (m\omega^2/2)\hat{x}^2, \tag{2}$$

where $\hat{x}$ and $\hat{p}$ are self-adjoint operators obeying the usual canonical commutation relations. Then, Wang's supplementary condition (3.17) on the wave function becomes

$$\hat{\phi}_1\psi = \hat{H}_0\psi - i\hbar\gamma\psi[\ln\psi - \tfrac{1}{2}\ln(\psi^*\psi)] = 0, \tag{3}$$

with ψ obeying a linear Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = \hat{H}\psi, \tag{4}$$

 0022-2488/90/020348-03$03.00

where the explicit form of the Hamiltonian operator $\hat{H}$ will be irrelevant for our purposes.

To our knowledge, all phenomenological quantum models of the damped harmonic oscillator enjoy the natural property that as $\gamma \to 0$, any state vector ψ_γ reduces to a state vector ψ_0 of the undamped oscillator. This is assumed by Wang himself in so far as he considers the case of small damping, with $\delta = \gamma/(\omega^2 - \gamma^2/4)^{1/2} \ll 1$, and makes use of a perturbative approximation method by expanding the Hamiltonian operator and wave function in powers of the small parameter δ. To the first order in δ, one should have [see Wang's Eq. (3.5)]

$$\psi = \psi_0 + \delta\psi_1, \quad \psi_0, \psi_1 \in \mathscr{H}, \tag{5}$$

where ψ_0 obeys the Schrödinger equation

$$i\hbar \frac{\partial \psi_0}{\partial t} = \hat{H}_0 \psi_0. \tag{6}$$

In other words, ψ_0 is a state vector of the undamped oscillator.

Therefore, not being aware of any convincing argument to the contrary, we shall adopt the natural point of view that any admissible quantum theory of the damped harmonic oscillator must reproduce the ordinary theory of the quantum oscillator as the friction coefficient γ tends to zero. What we intend to prove in the sequel is that the quantization scheme devised by Wang does not meet this requirement, and for this reason must be dismissed as unacceptable. As a byproduct of our investigation, it will also be shown that Wang's theory can only be consistent for a restricted set of values of the friction coefficient γ, and this is a further fatal objection against it.

III. INCONSISTENCY OF THE THEORY

Let us define the nonlinear operator $\hat{\mathscr{O}}$ by

$$\hat{\mathscr{O}}\psi = \psi[\ln \psi - \tfrac{1}{2}\ln(\psi^*\psi)]. \tag{7}$$

In order to ascribe a precise meaning to $\hat{\mathscr{O}}$, and therefore to Eq. (3), we shall take the principal determination of the logarithm.[27] Thus, with

$$\psi = |\psi| e^{i\theta}, \quad -\pi < \theta \leqslant \pi, \tag{8}$$

we have

$$\hat{\mathscr{O}}\psi = i\theta\psi. \tag{9}$$

Also, the operator $\hat{\mathscr{O}}$ has a few interesting properties that we now explore. From Eqs. (8) and (9) it follows that

$$|(\hat{\mathscr{O}}\psi)(x)| = |\theta(x)|\,|\psi(x)| \leqslant \pi|\psi(x)|, \tag{10}$$

whence

$$\|\hat{\mathscr{O}}\psi\| \leqslant \pi\|\psi\|, \tag{11}$$

showing that $\hat{\mathscr{O}}$ is a bounded operator. This does not mean that $\hat{\mathscr{O}}$ is necessarily continuous, because it is not a linear operator. Notice further that

$$\hat{\mathscr{O}}(\lambda\psi) = \lambda\hat{\mathscr{O}}\psi, \quad \lambda > 0, \tag{12}$$

so that $\hat{\mathscr{O}}$ is a positive-homogeneous operator in spite of its nonlinear character. By rewriting Eq. (3) in the form

$$\hat{H}_0\psi - i\hbar\gamma\hat{\mathscr{O}}\psi = 0, \tag{13}$$

the linearity of $\hat{H}_0$ together with the positive-homogeneity of $\hat{\mathscr{O}}$ allows us to require that any solution $\psi_\gamma \in \mathscr{H}$ to Eq. (13) be chosen in such a way that $\|\psi_\gamma\| = 1$. Since the Schrödinger equation (4) conserves probability, we may always regard as normalized any simultaneous solution to Eqs. (3) and (4).

Before going forward to what we want to establish, it is necessary to consider an ancillary result.

Lemma: Let ψ be any normalized vector of the Hilbert space of states $\mathscr{H}$. Then

$$\|\hat{H}_0\psi\|^2 \geqslant (\hbar^2\omega^2/4). \tag{14}$$

Proof: Let $\{\varphi_n\}_{n=0}^{\infty}$ be the orthonormal basis of $\mathscr{H}$ made up with the eigenvectors φ_n of $\hat{H}_0$, which are such that

$$\hat{H}_0\varphi_n = (n + 1/2)\hbar\omega\varphi_n. \tag{15}$$

It is possible to write

$$\psi = \sum_{n=0}^{\infty} c_n\varphi_n, \quad c_n \in \mathbb{C}, \tag{16}$$

with

$$\|\psi\|^2 = \sum_{n=0}^{\infty} |c_n|^2 = 1. \tag{17}$$

Therefore,

$$\hat{H}_0\psi = \sum_{n=0}^{\infty} c_n\left(n + \frac{1}{2}\right)\hbar\omega\varphi_n, \tag{18}$$

so that

$$\begin{aligned}(\hat{H}_0\psi, \hat{H}_0\psi) &= \sum_{n=0}^{\infty} \hbar^2\omega^2\left(n + \frac{1}{2}\right)^2 |c_n|^2 \\ &\geqslant \frac{\hbar^2\omega^2}{4}\sum_{n=0}^{\infty}|c_n|^2 \\ &= \frac{\hbar^2\omega^2}{4},\end{aligned} \tag{19}$$

and the proof is complete.

We are now prepared to state and prove our main results.

Theorem: Let $\psi_0 \in \mathscr{H}$ be a normalized state vector of the usual harmonic oscillator. Then there exists no solution $\psi_\gamma \in \mathscr{H}$ to Wang's equations (4) and (13) such that $\psi_\gamma \to \psi_0$ as $\gamma \to 0$.

Proof: As we have previously remarked, any solution $\psi_\gamma \in \mathscr{H}$ to Eqs. (4) and (13) can be taken to be normalized. Accordingly, let us assume that $\|\psi_\gamma\| = 1$ and rewrite Eq. (13) in the form

$$\hat{\mathscr{O}}\psi_\gamma = (i\hbar\gamma)^{-1}\hat{H}_0\psi_\gamma. \tag{20}$$

This equation combined with Eq. (11) leads to

$$\|\psi_\gamma\|^2 \geqslant \pi^{-2}\|\hat{\mathscr{O}}\psi_\gamma\|^2 = (\pi^2\hbar^2\gamma^2)^{-1}\|\hat{H}_0\psi_\gamma\|^2. \tag{21}$$

Having recourse to the Lemma we conclude that

$$\|\psi_\gamma\|^2 \geqslant \frac{\omega^2}{4\pi^2\gamma^2} \xrightarrow[\gamma\to 0]{} \infty. \tag{22}$$

This contradicts the assumption that ψ_γ is normalized and, moreover, shows that ψ_γ does not converge to an element of $\mathscr{H}$ as $\gamma \to 0$. The proof is complete.

Corollary: There exists no nontrivial solution to Eq. (13) in $\mathscr{H}$ if $\gamma < \omega/2\pi$.

Proof: Let $\psi_\gamma \in \mathscr{H}$ be a nontrivial solution (not necessarily normalized) to Eq. (13). From Eqs. (21), (19), and (17) it follows at once that

$$\|\psi_\gamma\|^2 \geqslant (\omega^2/4\pi^2\gamma^2)\|\psi_\gamma\|^2, \tag{23}$$

hence

$$\gamma \geqslant \omega/2\pi. \tag{24}$$

This restriction on the allowed values of γ implies the rejection of Wang's theory as unphysical, and characterizes as meaningless the perturbative approximation scheme employed in his paper, since it presupposes the validity of the model for γ/ω arbitrarily small.

IV. CONCLUSION

For one-dimensional nonconservative systems it is not difficult to understand why any attempt along the lines suggested by Wang will inevitably fail. If one insists that one is dealing with a genuine Hamiltonian system, although constrained, it must be possible to solve the constraint equations and go over to a reduced phase space (x^*, p^*) endowed with a Hamiltonian $H^*(x^*, p^*)$ and where there are no constraints. Making use of the path-integral quantization method, for instance, the formula for the propagator in the reduced phase space can be expressed in terms of the original phase space at the expense of a modification of the integration measure.[28] In any case, if the original phase space has dimension $2N$, the reduced one has at most dimension $2N-2$, if there is only one first-class constraint. The situation considered by Wang corresponds to $N=1$, and in the most favourable circumstances one would end up with a zero- dimensional phase space. Of course, even at the classical level such a theory is devoid of physical content.

ACKNOWLEDGMENT

This work was partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq, Brazil.

[1] P. A. M. Dirac, *Lectures on Quantum Mechanics*, Belfer Graduate School of Science Monograph Series No. 2 (Yeshiva University, New York, 1964); K. Sundermeyer, *Constrained Dynamics*, Lecture Notes in Physics, Vol. 169 (Springer, New York, 1982).
[2] C. F. Hayes and J. M. Jankowski, Nuovo Cimento B **58**, 494 (1968).
[3] C. F. Hayes, J. Math. Phys. **10**, 1555 (1969).
[4] C. A. P. Galvão and N. A. Lemos, J. Math. Phys. **29**, 1588 (1988).
[5] H. Bateman, Phys. Rev. **38**, 815 (1931).
[6] E. Kanai, Progr. Theor. Phys. **3**, 440 (1948).
[7] D. M. Greenberger, J. Math. Phys. **20**, 762 (1979).
[8] J. R. Ray, Am. J. Phys. **47**, 626 (1979).
[9] D. M. Greenberger, J. Math. Phys. **20**, 771 (1979).
[10] M. D. Kostin, J. Chem. Phys. **57**, 3589 (1972).
[11] R. W. Hasse, J. Math. Phys. **16**, 2005 (1975).
[12] B. K. Skagerstam, J. Math. Phys. **18**, 308 (1977).
[13] W. Stocker and K. Albrecht, Ann. Phys. (NY) **117**, 436 (1979).
[14] D. Schuch, K. M. Chung, and H. Hartmann, J. Math. Phys. **24**, 1652 (1983); **25**, 3086 (1984).
[15] L. Brül and H. Lange, J. Math. Phys. **25**, 786 (1984).
[16] I. R. Svin'in, Teor. Mat. Fiz. **27**, 270 (1976) [English translation: Theor. Math. Phys. **27**, 478 (1976)].
[17] J. Messer, Acta Phys. Austriaca **50**, 75 (1979).
[18] J. R. Brinati and S. S. Mizrahi, J. Math. Phys. **21**, 2154 (1980).
[19] H. Dekker, Phys. Rep. **80**, 1 (1981).
[20] W. H. Brittin, Phys. Rev. **77**, 396 (1950). Brittin's results have been partially corrected by Lemos (Ref. 24).
[21] P. Havas, Bull. Am. Phys. Soc. **1**, 337 (1956).
[22] J. Messer, Lett. Math. Phys. **2**, 281 (1978).
[23] I. K. Edwards, Am. J. Phys. **47**, 153 (1979).
[24] N. A. Lemos, Phys. Rev. D **24**, 2338 (1981).
[25] N. A. Lemos, Am. J. Phys. **49**, 1181 (1981).
[26] Y. Wang, J. Phys. A: Math. Gen. **20**, 4745 (1987).
[27] A. Markushevich, *Teoría de las Funciones Analíticas* (Mir, Moscow, 1970), Vol. I, pp. 18 and 187, 188.
[28] L. D. Faddeev, Teor. Mat. Fiz. **1**, 3 (1969) [English translation: Theor. Math. Phys. **1**, 1 (1970)].

Novel correspondence of the fixed-seniority and the fixed-isospin averages in their reduction formulas

Masao Nomura
Institute of Physics, College of Arts and Sciences, University of Tokyo, Komaba, Tokyo 153, Japan

(Received 22 July 1988; accepted for publication 20 September 1989)

It is shown that known reduction formulas for partial averages of two different kinds, the fixed-seniority and the fixed-isospin averages, are transcribed into each other by replacement of arguments. The average here is that of a general k-body operator in a finite space of fermions (bosons). Description of the formulas in a common form in terms of eigenvalues of Casimir operators is also exploited.

I. INTRODUCTION

Reduction formulas have been deduced for kinds of partial averages of a general k-body operator $O(k)$ in the space of n fermions (or bosons) being distributed over a finite set of orbits.[1–9] The partial average means the average under fixing of one or a few quantum numbers λ that partition the model space. Whether the reduction formula exists or not depends on λ.[4]

In the case where the partial average is characterized with simple propagations,[4] the reduction relation has the form[1–9]

$$\sum_\alpha \frac{\langle n\lambda\alpha|O(k)|n\lambda\alpha\rangle}{\dim(n,\lambda)} = \sum_{\lambda'} Z(n\lambda,k\lambda';\Omega)\sum_{\alpha'}\frac{\langle k\lambda'\alpha'|O(k)|k\lambda'\alpha'\rangle}{\dim(k,\lambda')}. \tag{1}$$

Here we assume that the n-body space is spanned by the orthonormalized states $\{|n\lambda\alpha\rangle\}$, where α denotes the set of additional quantum numbers. The symbol $\dim(n,\lambda)$ stands for the dimensionality of the states with the same n and λ. We denote by Ω half the $\dim(n=1,\lambda)$: $\Omega=(2j+1)/2$ in a single-j model of identical particles. We call Z the propagation coefficient (PC).[1–9] It is a polynomial in n and λ, and does not rely on orbital specifications such as the unitary group approach[10,11] and the prevalent shell model. Explicit forms of Z have been obtained for λ equal to the seniority v of identical particles,[2,3] and for λ equal to the quantum number of $U(N)$, $[\lambda_1,\lambda_2,\ldots,\lambda_N]$,[9] which in the case of $N=2$ means $\lambda=$ isospin T (or intrinsic spin S).

We often reexpress (1) with $\lambda=T$ and T_z as[1,2,7]

$$\sum_\alpha \frac{\langle nT\alpha|\,|O(k,r)|\,|nT\alpha\rangle}{\dim(n,T)} = \sum_r R(nT,kt;r)\sum_{\alpha'}\frac{\langle kt\alpha'|\,|O(k,r)|\,|kt\alpha'\rangle}{\dim(k,t)}. \tag{2}$$

Here $O(k,r)$ is the irreducible isotensor of rank r comprised of $O(k)$: $O(k)=\Sigma_r O(k,r)$. The factor R is a kind of PC, which is readily transformed into the PC $Z(nTT_z,ktt_z;\Omega)$.[7]

In the present paper we point out that the PC for the fixed-seniority average of identical fermions, $Z(nv)$, is transcribed into $R(nT)$ by the simple but unusual replacement of arguments $\Omega\leftrightarrow -r-2$, etc. Presence of this correspondence betwen PC's is not very obvious even if similarity exists between the quasispin formalism[12] for the seniority scheme and the isospin formalism. We show also that $Z(nv)$ and $R(nT)$ are rewritten in a common form in terms of eigenvalues of Casimir operators.

II. TRANSCRIPTION BETWEEN PROPAGATION COEFFICIENTS $Z(nv)$ AND $R(nT)$

Here we transcribe the PC for the fixed-seniority average of identical fermions into the PC for the fixed-isospin average.

The PC $Z(nv)$ is explicitly written as[2]

$$Z(nv)=\frac{((n-v)/2)!(\Omega-(k+v')/2)!(\Omega-v'+1)}{(n-k)!((k-v')/2)!(\Omega-(n+v)/2)!}Z'(nk) \tag{3a}$$

with

$$Z'(nk)=\sum_p\frac{(-1)^p(n-k+v'-2p)!(\Omega-(n+v-k+v')/2+p)!}{p!\,(v'-2p)!\,(\Omega-v'+1+p)!((n-k+v'-v)/2-p)!}, \tag{3b}$$

where the sum over p is taken such that none of the factorials could have a negative argument. The PC for the fixed-isospin average is also known as[1,2,7]

$$R(nT)=\frac{(2t+1)(n/2-T)!(n/2+T+1)!}{(n-k)!\,(k/2-t)!\,(k/2+t+1)!}\left(\frac{(2t-r)!\,(2T+r+1)!}{(2T-r)!\,(2t+r+1)!}\right)^{1/2}R'(nk), \tag{4a}$$

in which

$$R'(nk) = \sum_p \frac{(-1)^p (2t-p)!\,(n+2t-r-k-2p)!}{p!\,(2t-r-2p)!\,((n-k)/2+t-T-p)!((n-k)/2+T+t+1-p)!}. \tag{4b}$$

Allowed values of the seniority with a definite n are

$$v = n, n-2, n-4, ..., 0 (\text{or } 1). \tag{5}$$

Allowed values of the isospin are

$$T = \frac{n}{2}, \frac{n}{2}-1, \frac{n}{2}-2, ..., \begin{cases} r/2, & \text{if } n-r = \text{even} \\ (r+1)/2, & \text{otherwise.} \end{cases} \tag{6}$$

The quantities $Z'(nk)$, (3b), and $R'(nk)$, (4b), depend on $n-k+v'$ and $n-k+2t$, respectively, rather than n itself. Notice also $R(nT)$ is free from Ω.

Deduction of (3), (4), and some other PC's specified by (1) owes to the following property. Let $G(n,\lambda)$ be the eigenvalue of a p-body Casimir operator as to the symmetry λ.[3] It holds, due to a property of coefficients of fractional parentage (CFP's),[7] that

$$\sum_{\lambda''} G(n-1,\lambda'') Z(n\lambda, n-1\lambda''; \Omega) = (n-p) G(n,\lambda). \tag{7}$$

This determines $Z(n\lambda, n-1\lambda'; \Omega)$. Using the property that $Z(n\lambda)$ obeys a Chapman–Kolmogorov equation,[6] we obtain $Z(n\lambda, k\lambda'; \Omega)$. The PC in (1) resembles Rota's incidence coefficient.[13]

We find that $Z'(nk)$, (3b), is transcribed into $R'(n'k')$, (4b), and vice versa with the following replacements of variables:

$$v \leftrightarrow 2T - r, \quad v' \leftrightarrow 2t - r, \quad n \leftrightarrow n' - r,$$
$$k \leftrightarrow k' - r, \quad \Omega \leftrightarrow -r-2. \tag{8}$$

Factorials of negative integers have been interpreted as usual:

$$(-p)!/(-p-q)! = (-1)^q [(p+q-1)!/(p-1)!] \tag{9}$$

where p and q are positive integers. The condition (5) is transcribed into (6) with (8). Applying (8) to $Z(nv)$ we obtain

$$Z(n-r, 2T-r, k-r, 2t-r; -r-2) = \left(\frac{(2T-r)!\,(2t+r+1)!}{(2t-r)!\,(2T+r+1)!}\right)^{1/2} R(nT,kt;r). \tag{10}$$

The eigenvalue of a Casimir operator in the seniority scheme is given by

$$G(n,v) = (n-v)(2\Omega - n - v + 2)/4. \tag{11}$$

We apply the transcription (8) to (11), and get

$$G(n,v) \leftrightarrow -2(n'/2 - T)(n'/2 + T + 1) = G'(n',T). \tag{12}$$

The factor $G'(n,T)$ is described in terms of $T(T+1)$ and free from r. Invariance of $G(n,v)$ with replacement of n by $2\Omega - n + 2$ corresponds to invariance of $G'(n,T)$ with replacement of n by $-n-2$.

The reduction relation (1) is valid for the fermion and the boson systems alike. The PC for the fermion system is transcribed into that for the boson system by the replacement of Ω by $-\Omega$.[14–16] It is shown that replacing the argument $-r-2$ in (10) by $r+2$ yields the corresponding relation for the boson system.

As is seen from (3), $Z(nv)$ with $v=0$ is expressed as a single term. We combine this with (10) to obtain, for $r = 2T$,

$$R(nT,kt,r=2T) = \frac{(n/2-T)!\,(n/2+T+1)!\,(2T)!\,(2t+1)}{(k/2-t)!\,(k/2+t+1)!((n-k)/2+T+t+1)!((n-k)/2+T-t)!} \times \left(\frac{(2T+2t+1)!}{(4T+1)!\,(2t-2T)!}\right)^{1/2}, \tag{13}$$

a simple form of R that we could hardly infer from (4). This expression, combined with Eqs. (55) and (63) of Ref. 7, yields an unusual relation for the Racah coefficient:

$$\sum_{t'} \frac{2t'+1}{(p-t')!\,(p+t'+1)!} W(T,2T,t',t;Tt) = \frac{(2T)!}{(p+T+t+1)!\,(p+T-t)!} \times \left(\frac{(2T+2t+1)!}{(4T+1)!\,(2t-2T)!}\right)^{1/2}. \tag{14}$$

III. PROPAGATION COEFFICIENTS $Z(nv)$ AND $R(nT)$ IN A COMMON FORM

The objective of this section is to describe $Z(nv)$ and $R(nT)$ in a common form in terms of eigenvalues of Casimir operators.

A general PC $Z(n\lambda)$ in (1) has the following properties:[1–6]

(a) a polynomial of degree k in n;

(b) a polynomial of degree $[k/2]$ (the largest integer contained in $k/2$) in the eigenvalue of a bilinear Casimir operator attached to λ;

(c) $Z(n\lambda, k\lambda'; \Omega) = 0$ in the case $n < k$; and

(d) $Z(k\lambda, k\lambda'; \Omega) = \delta(\lambda,\lambda')$.

We easily check that $Z(nv)$ actually satisfies (a)–(d). In the case of $R(nT)$, (4), the statement (d) is modified to

$$(d') \quad R(kT,kt;\gamma) = \begin{cases} \delta(T,t), & \text{if } T \geqslant r/2, \\ 0, & \text{otherwise.} \end{cases}$$

The statements (a)–(d) for $Z(nv)$ still hold even if $G(n,v)$ is replaced by any one of $G(n-2,v)$, $G(n-4,v)$, etc., because of the property $G(n-a,v)$

$= G(n,v) + a(n - a/2 - \Omega - 1)/2$. Similarly, use of $G'(n-2,T)$, $G'(n-4,T)$, etc., in place of $G'(n,T)$ does not modify any one of (a)–(c) or (d′). Let us determine $Z(nv)$ explicitly using (a)–(d) only. We use (b) to expand $Z(nv)$ as

$$Z(nv) = A_0 + A_1 G(n,v) + A_2 G(n,v) G(n-2,v) + \cdots , \tag{15}$$

where the A_p's are coefficients independent of v. From (a) it follows that A_p is a polynomial of degree $n - 2p$ in n. Let us make use of (c) to determine the n dependence of the A's. Putting $n = v = 0$ in (15), we find that A_0 is a multiple of n. Here we have used the property $G(v,v) = 0$. Similarly, vanishing of $Z(nv)$ in the cases of $n = v = 1$ and of $n = 2$ and $v = 1$ means that A_0 can be divided by $n - 1$ and $n - 2$, respectively. Form the vanishing of $Z(v,v)$ with $v = 2$ it follows that we can divide A_1 by $n - 2$. Repeating this yields

$$A_p = C_p \binom{n-2p}{k-2p}, \tag{16}$$

where C_p is the coefficient independent of n and v. Let us use (d) to determine C_p. We put $n = k = v$ in the set of equations (15) and (16) to get $C_0 = \delta(k,v')$. Similarly, we get $C_0 + C_1 G(k,k-2) = \delta(k-2,v')$ from (15) with $n = k = v + 2$. Repeating this we get a set of equations that determines all the C_p's. We then obtain

$$Z(nv) = \sum_{p=0}^{[v'/2]} D_p \binom{n + v' - k - 2p}{v' - 2p} \times \prod_{q=1}^{(k-v')/2+p} G(n - 2q + 2, v), \tag{17a}$$

in which D_p means $C_{p+(k-v')/2}$ and is expressed as

$$D_p = (-1)^p \frac{(\Omega - (k+v')/2)!\,(\Omega + 1 - v')}{p!\,((k-v')/2)!\,(\Omega + p + 1 - v')!}. \tag{17b}$$

The expression obtained with $k = 6$ and $v' = 4$, for example, reads

$$Z(nv,6,4;\Omega) = \binom{n-2}{4} \frac{G(n,v)}{\Omega - 4} - \binom{n-4}{2} \frac{G(n,v)G(n-2,v)}{(\Omega-2)(\Omega-4)} + \frac{G(n,v)G(n-2,v)G(n-4,v)}{2(\Omega-1)(\Omega-2)(\Omega-4)}. \tag{18}$$

The same method is used to determine $R(nT)$. The statements (a)–(c) are common to $Z(nv)$ and $R(nT)$. The resultant expression of $R(nT)$ is just (17a) with G in it being replaced by G', (12), and with D_p being replaced by D'_p as follows:

$$D'_p = (-1)^{k/2-t} \frac{(2t+1)(2t-p)!}{p!\,(k/2-t)!\,(k/2+t+1)!}, \tag{19}$$

a factor that is related to D_p by means of (8). We thus describe $Z(nv)$ and $R(nT)$ in a unified way in terms of eigenvalues of Casimir operators.

[1]J. B. French, in *Isospin in Nuclear Physics*, edited by D. H. Wilkinson (North-Holland, Amsterdam, 1969), p. 259.
[2]M. Nomura, Prog. Theor. Phys. **51**, 489 (1974).
[3]J. C. Parikh, *Group Symmetries in Nuclear Structure* (Plenum, New York, 1978).
[4]C. Quesne, in *Theory and Applications of Moment Methods in Many-Fermion Systems*, edited by B. J. Dalton, S. M. Grimes, J. P. Vary, and S. A. Williams (Plenum, New York, 1980), p. 257.
[5]V. K. B. Kota, Ann. Phys. (NY) **134**, 221 (1981).
[6]M. Nomura, J. Math. Phys. **26**, 965 (1985).
[7]M. Nomura, J. Math. Phys. **27**, 536 (1986).
[8]M. Nomura, J. Math. Phys. **28**, 128 (1987).
[9]M. Nomura, Phys. Lett. **126**, 1 (1987).
[10]J. Paldus, J. Chem. Phys. **61**, 5321 (1974).
[11]M. Nomura, Phys. Rev. A **37**, 2709 (1988).
[12]R. D. Lawson and M. H. Macfarlane, Nucl. Phys. **66**, 80 (1965).
[13]G.-C. Rota, P. Doubilet, and R. Stanley, in *Finite Operator Calculus*, edited by G.-C. Rota (Academic, New York, 1975), p. 83.
[14]M. Nomura, Nucl. Phys. A **150**, 193 (1970).
[15]M. Nomura, Ann. Phys. (NY) **91**, 83 (1975).
[16]V. K. B. Kota, J. Phys. (Paris) **40**, L579 (1979).

Geometry and uncertainty

Francis J. Narcowich
Department of Mathematics, Texas A&M University, College Station, Texas 77843-3368

(Received 31 May 1989; accepted for publication 20 September 1989)

In the context of the Wigner–Weyl phase space formulation of quantum mechanics, a version of the uncertainty relations invariant under affine canonical transformations is derived. For a fixed Wigner distribution function possessing a finite covariance, "directions" of minimal uncertainty are found. The geometry of the Wigner ellipsoid and its Legendre transform, the dual Wigner ellipsoid, both of which are associated with the same covariance, is discussed. The results obtained are generalizations of the well-known fact that, for one degree of freedom, the area of the Wigner ellipse must be of order $\hbar$ or larger. Instead of area, which is an invariant only when $n = 1$, these results involve Poincaré invariants of certain curves and surfaces.

I. INTRODUCTION

In his review article on the semiclassical evolution of wave packets, Littlejohn[1] dispels the old myth that the spreading of a free wave packet is a quantum mechanical effect. Employing Gaussian wave packets, he gives an example illustrating that the spread one gets can be entirely accounted for classically. Using the covariance matrix for a Wigner distribution function (WDF), he then shows that this result is true for arbitrary states. In the course of his discussion, a number of questions arise. Is there a *geometric* way of viewing the uncertainty relations? Is there a formulation of the uncertainty relations that is invariant under affine canonical transformations? Is there a way of finding a direction of *minimal* uncertainty for arbitrary states, whether they are Gaussian or non-Gaussian, pure or mixed? Answers to these questions would help in distinguishing classical effects, such as spreading of wave packets, from truly quantum mechanical effects presently masked by classical ones.

The purpose of this paper is to answer these questions, and thus to explore the geometry of uncertainty. The paper is organized as follows. In the remainder of Sec. I, we will introduce notation and conventions basic to the work. In Sec. II, we will recall the definition of the covariance matrix for a WDF, and then give necessary and sufficient conditions for a matrix to be the covariance of a WDF. We will follow this up in Sec. III with a discussion of how a covariance matrix for a WDF transforms under affine canonical transformations. In Sec. IV, we briefly discuss Williamson's normal form for a real, symmetric matrix when the matrix is positive definite. We use this to obtain results crucial in the succeeding parts of the paper. In V, we will apply the results we have gotten in earlier sections to obtain a version of the uncertainty relations that is invariant under affine canonical transformations. The section will conclude with a discussion of the "directions" of *minimal* uncertainty. In the final section of the paper, Sec. VI, we will discuss the geometry of two ellipsoids associated with the covariance of a WDF, the *Wigner ellipsoid* and its Legendre transformation, the *dual Wigner ellipsoid*. Our results are generalizations of the well-known fact that for one degree of freedom, the area of the Wigner ellipse must be of order $\hbar$ or larger. Instead of area, which is an invariant only when $n = 1$, these results involve *Poincaré invariants* of certain curves and surfaces.

We will use the Wigner–Weyl phase space formulation of quantum mechanics to obtain our results. We do so because it is in this formulation that classical and quantum concepts are best compared. One may find brief reviews of it in Littlejohn (see Ref. 1, Appendix B), Narcowich,[2] and Narcowich and O'Connell.[3] For a more extensive review, see Hillery *et al.*[4]

Notation and conventions associated with the Wigner–Weyl formulation differ from author to author. In this paper, we will use the following notation. For a quantum system with n spinless degrees of freedom, the phase space is $\mathbb{R}^n \times \mathbb{R}^n \approx \mathbb{R}^{2n}$; we will denote this space by Γ. We will let $z = (q_1, q_2, \ldots, q_n; p_1, p_2, \ldots, p_n)$ denote a point in Γ. The convention that we will use in dealing with components is the same as that used by Littlejohn[1]; latin subscripts will run from 1 to n, and greek subscripts will run from 1 to $2n$. Also, we choose the units of both the p's and q's to be those of $\sqrt{\hbar}$. For z, the components z_α corresponding to $\alpha = 1, \ldots, n$ are $q_1, \ldots, q_n$, and those corresponding to $\alpha = n+1, \ldots, 2n$ are $p_1, \ldots, p_n$. We will take

$$\sigma(z,z') = \sum_{j=1}^{n} q_j' p_j - q_j p_j' \qquad (1.1)$$

to be the usual symplectic form on phase space. This can be written in the form

$$\sigma(z,z') = z'^T J z, \quad \text{where } J = \begin{pmatrix} 0_n & I_n \\ -I_n & 0_n \end{pmatrix}. \qquad (1.1')$$

Here 0_n and I_n are the $n \times n$ zero and identity matrices, respectively. When matrix notation is used, z is to be thought of as a column vector. Finally, the superscript "T" indicates the transpose of a matrix or vector.

If f is in $\mathscr{S}$, Schwartz space, then the *symplectic Fourier transform* of f is

$$\tilde{f}(a) \overset{\text{def}}{=} \int_\Gamma f(z) e^{i\sigma(a,z)}\, dz, \qquad (1.2)$$

where $a = (u_1, u_2, \ldots, u_n; v_1, v_2, \ldots, v_n)$ is a point in Γ', the dual of phase space, and

$$dz = \prod_{j=1}^{n} dq_j\, dp_j$$

is the standard Liouville measure on Γ. (With the obvious

 0022-2488/90/020354-11$03.00

differences accounted for, our conventions regarding the components of z apply to a. We will also denote the Liouville measure on Γ' by da. The units of the components of a are those of $1/\sqrt{\hbar}$.) The symplectic Fourier transform of f is directly related to the ordinary Fourier transform of f (see Ref. 3). One can invert (1.2) to get

$$f(z) = (2\pi)^{-2n} \int_{\Gamma'} \tilde{f}(a)\exp[i\sigma(z,a)]da. \qquad (1.3)$$

In the Wigner–Weyl formulation, both pure and mixed quantum states are represented by *Wigner distribution functions* (WDF's). Using the KLM conditions,[3,5–8] one can characterize WDF's in terms of their symplectic Fourier transforms. To state the KLM conditions, we begin by defining functions of $\hbar$-positive type. We say that a continuous function $F(a)$ defined on Γ' is of *$\hbar$-positive type* if, for every finite set of points $\{a_1,...,a_m\} \subset \Gamma'$, the Hermitian $m \times m$ matrix with (j,k) entry

$$F(a_j - a_k)\exp[i\hbar\sigma(a_k,a_j)/2] \qquad (1.4)$$

is non-negative. What the KLM conditions state is that $\tilde{\rho}(a)$ is the symplectic Fourier transform of a WDF $\rho(z)$ if and only if (i) $\tilde{\rho}$ is continuous and of $\hbar$-positive type and (ii) $\tilde{\rho}(0) = 1$. (The second condition merely ensures the correct normalization for the WDF.)

II. THE COVARIANCE MATRIX FOR A WDF

The standard p-q uncertainty relations are actually constraints on the second-order moments of a WDF[3,9,10]; that is, they are constraints on the covariance matrix for a WDF. When the uncertainty relations are stated in the usual way, they are *not* invariant under symplectic changes of coordinates[1] and they do not behave in any "nice" way under such changes of coordinates. The situation is considerably different for the covariance matrix itself (see Ref. 1, § 7.2). In this section, we will discuss necessary and sufficient conditions for a $2n \times 2n$ matrix to be the covariance matrix for a WDF. In the next section, we will review the behavior of the covariance under general affine transformations. In doing so, without using the metaplectic group, we will recover the result (see Ref. 1, § 6.3) that the set of WDF's is invariant under the affine canonical transformation's covariance matrix. We will begin with the following result.

Lemma 2.1: Let $F(a)$ be a continuous function defined on Γ'. Suppose that F is twice continuously differentiable near $a = 0$. Denote the Hessian of F at a by $F''(a)$. If F is of $\hbar$-positive type, then the matrix

$$-F''(0) + (i\hbar/2)J \qquad (2.1)$$

is Hermitian and non-negative.

Proof: Because F is of $\hbar$-positive type, the matrix given in (1.4) is non-negative. Consequently, if the points in Γ' are chosen to be $\epsilon a_1,...,\epsilon a_m$, where ϵ is an arbitrary real number, and if $\lambda_1,...,\lambda_m$ are arbitrary complex numbers, then

$$\mathscr{F}(\epsilon) = \sum_{j=1,k=1}^{m} \bar{\lambda}_j\lambda_k F(\epsilon(a_j - a_k)) \times \exp[i\epsilon^2\hbar\sigma(a_k,a_j)/2] \geqslant 0. \qquad (2.2)$$

Choose the λ's in (2.2) so that

$$\sum_{j=1}^{m}\lambda_j = 0, \qquad (2.3)$$

and then fix them. Also, fix the a_j's. Because F is twice continuously differentiable for all a sufficiently close to $a = 0$, we see that $\mathscr{F}$ is also twice continuously differentiable near $\epsilon = 0$. Using (2.2) and (2.3), we find that

$$\mathscr{F}(0) = F(0)\left|\sum_{k=1}^{m}\lambda_k\right|^2 = 0. \qquad (2.4)$$

Since $\mathscr{F}(\epsilon)$ is non-negative, (2.4) implies that $\mathscr{F}$ has a minimum at $\epsilon = 0$. By the second derivative test,

$$\mathscr{F}''(0) \geqslant 0. \qquad (2.5)$$

A straightforward calculation gives us

$$\mathscr{F}''(0) = \sum_{j=1,k=1}^{m} \bar{\lambda}_j\lambda_k[(a_j - a_k)^T F''(0)(a_j - a_k) + i\hbar\sigma(a_k,a_j)] \geqslant 0. \qquad (2.6)$$

We can simplify the expression in (2.6) by using (2.3), for terms of the form

$$\sum_{j=1,k=1}^{m} \bar{\lambda}_j\lambda_k[a_j^T F''(0)a_j] = \left(\sum_{k=1}^{m}\lambda_k\right)(\cdots) = 0. \qquad (2.7)$$

Using this fact and (1.1′), we get

$$\mathscr{F}''(0) = \sum_{j=1,k=1}^{m} \bar{\lambda}_j\lambda_k[-a_k^T F''(0)a_j - a_j^T F''(0)a_k + i\hbar a_j^T J a_k] \geqslant 0. \qquad (2.8)$$

We may rewrite this last expression as

$$\mathscr{F}''(0) = \xi^\dagger[-F''(0) - F''(0)^\dagger + i\hbar J]\xi \geqslant 0,$$

$$\text{where } \xi = \sum_{k=1}^{m}\lambda_k a_k. \qquad (2.9)$$

(Here, "†" is the transpose conjugate of a vector or a matrix.) Now, let $m = 3$ and choose $\lambda_1 = 1$, $\lambda_2 = i$, $\lambda_3 = -1 - i$, and $a_3 = 0$. With these choices made, $\xi = a_1 + ia_2$ is an arbitrary vector in $\mathbb{C}^{2n}$, and so from (2.9) we have that

$$-F''(0) - F''(0)^\dagger + i\hbar J \qquad (2.10)$$

is non-negative.

The final step is to observe that F being of $\hbar$-positive type implies that $F(a) = \bar{F}(-a)$, which in turn forces $F''(0)$ to be a real, Hermitian matrix. This observation, (2.10), and the fact that iJ is Hermitian immediately yield that the matrix in (2.1) is both Hermitian and non-negative. ■

Let $\|z\|$ be the Euclidean norm on Γ, and suppose that $\rho(z)$ is a WDF for which $(\|z\|^2 + 1)\rho(z)$ is in $L^1(\Gamma)$. This condition not only guarantees the finiteness of all of the moments up to and including those of order^2, but it also implies that $\tilde{\rho}(a)$ is twice continuously differentiable on all of Γ'. (Standard arguments from Fourier analysis may be used to prove this fact. We omit the details.) Suppose that z_0 is the expectation value for z with respect to ρ; that is,

$$z_0 \equiv \int_\Gamma z\rho(z)dz. \qquad (2.11)$$

Translating the argument of ρ by z_0 results in a new WDF:

$$\rho_0(z) \equiv \rho_0(z+z_0). \tag{2.12}$$

Obviously, the expectation value of z with respect to ρ_0 is 0. In addition, when taken with respect to ρ_0, the expectation values of quantities such as q_1^2 and p_1^2 turn out to be Δq_1^2 and Δp_1^2 for the original state ρ. Put differently, the covariance matrix for ρ, which is defined to be the $2n\times 2n$ matrix C with (α,β) entry,

$$C_{\alpha,\beta} \equiv \int_\Gamma (z-z_0)_\alpha (z-z_0)_\beta \rho(z)dz, \tag{2.13}$$

is also the covariance matrix for ρ_0. To see this, note that since the expectation value of z with respect to ρ_0 is 0, the covariance matrix for ρ_0 has

$$\int_\Gamma z_\alpha z_\beta \rho_0(z)dz, \tag{2.14}$$

for its (α,β) entry. A change of variables $z\to z-z_0$ in (2.14) turns the expression there into the right side of (2.13). Thus, as claimed, the covariance matrices for the two WDF's are equal.

There is an obvious but important connection between the covariance matrix C and the Hessian $\tilde\rho_0''$. Because we have assumed that $(\|z\|^2+1)\rho(z)$ is in $L^1(\Gamma)$, we have that the same is true for ρ_0. This implies that $z_\alpha z_\beta \rho_0(z)$ is also in $L^1(\Gamma)$, that $\tilde\rho_0(a)$ is a twice continuously differentiable function, and that the components of $\tilde\rho''$ may be computed by differentiating under the integral sign used in $\tilde\rho_0$. Doing this differentiation gives

$$(\tilde\rho_0''(0))_{\alpha,\beta} = -\sum_{\alpha'=1,\beta'=1}^{2n} J_{\alpha,\alpha'} J_{\beta,\beta'} \int_\Gamma z_{\alpha'} z_{\beta'} \rho_0(z)dz. \tag{2.15}$$

Using the components of the covariance matrix C in (2.15), we get

$$(\tilde\rho_0''(0))_{\alpha\beta} = -\sum_{\alpha'=1,\beta'=1}^{2n} J_{\alpha,\alpha'} J_{\beta,\beta'} C_{\alpha',\beta'}. \tag{2.16}$$

Using (2.16) and the definition of J, we see that

$$\tilde\rho_0''(0) = -JCJ^T. \tag{2.17}$$

Finally, because ρ_0 is a WDF, $\tilde\rho_0$ satisfies the KLM conditions and is of $\hbar$-positive type. As we have already noted, $\tilde\rho_0$ is, by virtue of our assumptions on ρ, twice continuously differentiable on Γ'. Hence, Lemma 2.1 applies, and $-\tilde\rho_0''(0)+(i\hbar/2)J$ is a Hermitian, non-negative matrix. Replacing the Hessian in this matrix by the right side of (2.17), we find that the Hermitian matrix

$$JCJ^T + (i\hbar/2)J \tag{2.18}$$

is non-negative. Multiplying (2.18) on the right by J and on the left by $J^\dagger = J^T = -J$ and noting that $J^2 = -I$, we arrive at this result.

Theorem 2.2: If ρ is a WDF for which $(\|z\|^2+1)\rho(z)$ is in $L^1(\Gamma)$, then the covariance matrix C defined by (2.13) exists and the matrix

$$C + (i\hbar/2)J \tag{2.19}$$

is Hermitian and non-negative.

For Gaussian WDF's and coherent states in optics, the condition (2.19) is known; Narcowich,[11] Narcowich and O'Connell,[12] and Yuen[13] state it explicitly, and Simon *et al.*[14] formulate it implicitly. It also appears in disguised form in a much earlier work by Lindblad,[15] who discusses it in the context of expectations of the Weyl operators resulting in Gaussians.

The usual p-q uncertainty relations are an easy consequence of this theorem.[3,16] To see this, note that the non-negativity and Hermiticity of (2.19) imply that the submatrix

$$\begin{pmatrix} \Delta q_j^2 & C_{j,j+n} + i\hbar/2 \\ C_{j+n,j} - i\hbar/2 & \Delta p_j^2 \end{pmatrix} \tag{2.20}$$

is Hermitian and non-negative. The usual conditions for a Hermitian matrix to be non-negative then imply that

$$\Delta q_j^2 \Delta p_j^2 - |C_{j,j+n}|^2 - \hbar^2/4 \geqslant 0, \tag{2.21}$$

from which the p-q uncertainty relations

$$\Delta q_j \Delta p_j \geqslant \hbar/2 \tag{2.22}$$

follow at once.

The difficulty with the uncertainty relations as stated in (2.22) is that the individual products $\Delta q_j \Delta p_j$ are *not* invariant under linear canonical transformations, and so they do not provide a truly good way of distinguishing the effects of classical mechanics from those of quantum mechanics. A way around this difficulty is to first notice that (2.19) *is* invariant under such transformations, and then to give *invariant* conditions equivalent to the matrix (2.19) being non-negative. This we will do later. For now, we will conclude this section by showing that the conditions on C in Theorem 2.2. are actually both necessary and sufficient for C to be a covariance matrix for a WDF. We will first prove the technical lemma below.

Lemma 2.3: If C is any real, symmetric $2n\times 2n$ matrix for which the matrix $C+i\eta J$ is non-negative for some real $\eta\neq 0$, then C must be positive definite.

Proof: We begin by showing that C is non-negative. If not, then C has an eigenvalue that is negative. Suppose that it has $\lambda<0$ as an eigenvalue. Because C is real and symmetric, we may choose an eigenvector X_- corresponding to λ so that X_- is real. Then, we have that

$$X_-^\dagger (C+i\eta J)X_- = \lambda \|X_-\|^2 + i\eta X_-^\dagger J X_-.$$

Since X_- is real, we also have that

$$X_-^\dagger J X_- = X_-^T J X_- = \sigma(X_-,X_-) = 0.$$

Thus our last equation becomes

$$X_-^\dagger (C+i\eta J)X_- = \lambda \|X_-\|^2.$$

Since $\lambda<0$, this implies that $X_-^\dagger (C+i\eta J)X_- < 0$, which contradicts the non-negativity of $C+i\eta J$. Hence C has no negative eigenvalues and is therefore a non-negative matrix.

Suppose that C has 0 as an eigenvalue. Because C is real, we may again choose an eigenvector X_0 corresponding to 0 so that X_0 is real. Define the vector $X_\epsilon \equiv (I+i\epsilon J)X_0$. Observe that, since X_0 is real, $X_0^T J X_0 = 0$, and since X_0 is an eigenvector of C corresponding to the eigenvalue 0, we have $X_0^T C = 0$ as well as $CX_0 = 0$. This allow us to make the following calculation:

$$X_\epsilon^\dagger (C+i\eta J)X_\epsilon = X_0^T(I+i\epsilon J)(C+i\eta J)(I+i\epsilon J)X_0$$
$$= X_0^T[C+i\epsilon JC+i\epsilon CJ-\epsilon^2 JCJ]X_0 + i\eta X_0^T J[I+2i\epsilon J+\epsilon^2 I]X_0$$
$$= \epsilon^2 (JX_0)^T C(JX_0) + \epsilon\eta\|X_0\|^2 = \epsilon\eta\|X_0\|^2[1+O(\epsilon)]. \tag{2.23}$$

If we now choose ϵ to be small and opposite in sign to η, we get that

$$X_\epsilon^\dagger (C+i\eta J)X_\epsilon < 0, \tag{2.24}$$

which again contradicts the non-negativity of $C+i\eta J$. Hence C cannot have 0 as an eigenvalue and so must be positive definite. ■

The point of this lemma is that it shows that if a real, symmetric matrix C satisfies $C+(i\hbar/2)J$ being non-negative, then C is positive definite and the Gaussian function

$$\rho(z) \overset{\text{def}}{=} (2\pi)^{-n}(\det C)^{-1/2}e^{-(1/2)z^T C^{-1} z} \tag{2.25}$$

is a WDF with covariance C. See Ref. 11, § III. This observation shows that the result below is true.

Theorem 2.4: A real, symmetric $2n\times 2n$ matrix C is the covariance matrix for some WDF if and only if the matrix in (2.19) is non-negative.

III. INVARIANCE

We now want to review the behavior of the covariance of a WDF under affine canonical transformations. Before we do that, we should review what it means for a *quantum mechanical* state to be subjected to what is, after all, a *classical* transformation.

A $2n\times 2n$ matrix S is symplectic if and only if S satisfies

$$\sigma(Sz,Sz') = \sigma(z,z'), \tag{3.1}$$

for every pair of points z, z' in phase space Γ. [The symplectic form σ was defined in (1.1).] Two equivalent matrix forms of this are

$$S^T JS = J \quad \text{and} \quad SJS^T = J. \tag{3.1$'$}$$

An *affine* transformation $A_c: \Gamma\to\Gamma$ is said to be *canonical* if it has the form

$$A_c z = Sz + \zeta, \tag{3.2}$$

where S is a symplectic matrix and $\zeta\in\Gamma$. Let $f(z)$ be a function defined on Γ. We follow Littlejohn (see Ref. 1, § 6.3) in defining

$$M_{cl} f(z) \overset{\text{def}}{=} f(A_c^{-1} z) = f(S^{-1}(z-\zeta)). \tag{3.3}$$

In Ref. 1, Littlejohn uses the action of the metaplectic group on wave functions to show that M_{cl} takes WDF's into WDF's. One can also give a proof using the KLM conditions. Since this proof is short and self-contained, we will present it here.

Theorem 3.1: The operator M_{cl} defined in (3.3) transforms WDF's into WDF's.

Proof: Let ρ be a WDF and set $\rho'(z)\equiv M_{cl}\rho(z)$. A straightforward computation using (1.1), (1.2), (3.3), and well-known properties of symplectic matrices gives us that the symplectic Fourier transform on ρ' is given by

$$\tilde\rho'(a) = e^{i\sigma(S^{-1}a,S^{-1}\zeta)}\tilde\rho(S^{-1}a). \tag{3.4}$$

Since ρ is a WDF, $\tilde\rho$ obeys the KLM conditions. Thus it is continuous, of $\hbar$-positive type, and satisfies $\tilde\rho(0)=1$. It is obvious that $\tilde\rho'$ is also continuous and satisfies $\tilde\rho'(0)=1$. To show that ρ' is a state, we need only show that $\tilde\rho'$ is of $\hbar$-positive type. This amounts to showing the non-negativity of the $m\times m$ matrix

$$K'_{jk} = \tilde\rho'(a_j-a_k)\exp[i\hbar\sigma(a_k,a_j)/2], \tag{3.5}$$

where $\{a_1,\dots,a_m\}$ is an arbitrary finite subset in the dual of phase space, Γ'.

We see from (3.1), (3.4), and (3.5) that

$$\begin{aligned} K'_{jk} &= e^{i\sigma(b_j-b_k,S^{-1}\zeta)}\tilde\rho(b_j-b_k)e^{i(\hbar/2)\sigma(a_k,a_j)} \\ &= e^{i\sigma(b_j-b_k,S^{-1}\zeta)}\tilde\rho(b_j-b_k)e^{i(\hbar/2)\sigma(S^{-1}a_k,S^{-1}a_j)} \\ &= e^{i\sigma(b_j-b_k,S^{-1}\zeta)}\tilde\rho(b_j-b_k)e^{i(\hbar/2)\sigma(b_k,b_j)} \end{aligned} \tag{3.6}$$

where $b_j = S^{-1}a_j$. If we set

$$K_{jk} = \tilde\rho(b_j-b_k)e^{i(\hbar/2)\sigma(b_k,b_j)}, \tag{3.7}$$

then, because $\tilde\rho$ is of $\hbar$-positive type, we see that K_{jk} is a non-negative Hermitian matrix. We may rewrite (3.6) as

$$K'_{jk} = e^{i\sigma(b_j-b_k,S^{-1}\zeta)}K_{jk}. \tag{3.8}$$

From (3.8), it is obvious that K'_{jk} is Hermitian. To see that it is non-negative, let $\lambda'_1,\dots,\lambda'_m$ be arbitrary complex numbers. Note that the non-negativity of K_{jk} yields this:

$$\sum_{j,k=1}^m \bar\lambda'_j\lambda'_k K'_{jk} = \sum_{j,k=1}^m \bar\lambda'_j\lambda'_k e^{i\sigma(b_j-b_k,S^{-1}\zeta)}K_{jk}$$
$$= \sum_{j,k=1}^m \bar\lambda_j\lambda_k K_{jk} \geqslant 0, \tag{3.9}$$

where $\lambda_k = \lambda'_k\exp[-i\sigma(b_k,S^{-1}\zeta)]$. Thus we have that $\tilde\rho'$ is of $\hbar$-positive type and so ρ' is a WDF. ■

The next question that we want to address is, "How does the covariance matrix for a WDF ρ transform when we replace ρ by $\rho'\equiv M_{cl}\rho$?" Using matrix notation, the covariance matrices for ρ and ρ' are, respectively,

$$C = \int_\Gamma (z-z_0)(z-z_0)^T\rho(z)dz$$

and (3.10)

$$C' = \int_\Gamma (z-z_0')(z-z_0')^T\rho'(z)dz.$$

Here, z_0 and z_0' are the expectation values of z relative to ρ and ρ'. Making the change of integration variables $z\to S^{-1}(z-\zeta)$ in the integral defining z_0' and using the fact that $\det S=1$ for a symplectic matrix, one can show that

$$z_0' = Sz_0 + \zeta = A_c z_0. \tag{3.11}$$

Substituting this into the expression for C' in (3.10) and making the same change of integration variables, one gets

$$C' = \int_\Gamma [S(z - z_0)][S(z - z_0)]^T \rho(z) dz,$$

which can be rewritten as

$$C' = SCS^T. \tag{3.12}$$

This result agrees with that of Littlejohn (see Ref. 1, § 7.1).

We also wish to point out that this is completely consistent with characterizing a covariance matrix in the way we did in Sec. II. There, we showed that a $2n \times 2n$ matrix is a covariance for a WDF if and only if it is real, symmetric, and has the property that $C + (i\hbar/2)J$ is non-negative. If C satisfies these conditions, then $C' = SCS^T$ is certainly real and symmetric. In addition, for every $W \in \mathbb{C}^{2n}$ the matrix C' also satisfies

$$W^\dagger(C' + (i\hbar/2)J)W = W^\dagger(SCS^T + (i\hbar/2)SJS^T)W$$
$$= (S^TW)^\dagger(C + (i\hbar/2)J)(S^TW) \geqslant 0,$$

and so $C' + (i\hbar/2)J$ is non-negative.

A computation nearly identical to the one we just made gives us that if $C' = SCS^T$, then

$$C' + i\eta J = S(C + i\eta J)S^T. \tag{3.13}$$

This is an equation that will prove to be useful in getting an invariant version of the uncertainty relations.

IV. A MATRICIAL INTERLUDE

We will now interrupt our discussion to give a brief discussion of topics that are related to Williamson's normal form of a matrix under symplectic transformation of coordinates.[13,14,17,18] Most of the complications that arise in connection with this form can be avoided by working with a positive definite matrix. Since covariance matrices for WDF's are positive definite, we will obtain all the results we need by restricting our attention to such matrices.

Let C be a real, symmetric $2n \times 2n$ matrix and let

$$\chi(\eta) \overset{\text{def}}{=} \det(C + i\eta J). \tag{4.1}$$

Because C is both real and symmetric, we can use elementary properties of determinants to show that if η is a root of $\chi(\eta) = 0$, then $\pm\eta$ and $\pm\bar{\eta}$ are also roots. If we also assume that C is a positive definite matrix, then we have that

$$C + i\eta J = C^{1/2}(I + i\eta C^{-1/2}JC^{-1/2})C^{1/2}, \tag{4.2}$$

which implies that

$$\chi(\eta) = \det(C)\det(I + i\eta C^{-1/2}JC^{-1/2}). \tag{4.3}$$

Thus the roots of $\chi(\eta) = 0$ are reciprocals of the eigenvalues of the invertible Hermitian matrix $-iC^{-1/2}JC^{-1/2}$, and so they are real and nonzero. Putting all this together results in the following.

Proposition 4.1: If C is a real, symmetric, positive definite $2n \times 2n$ matrix, then any root η of $\chi(\eta) = 0$ is real and nonzero; moreover, $-\eta$ is also a root. Finally, if S is a symplectic matrix, then the roots of $\chi(\eta) = 0$ are invariant under the transformation $C \to SCS^T$.

Proof: Only the statement about symplectic invariance of the roots requires comment. From (3.1′) and the fact that S being symplectic implies that $\det S = 1$, we have

$$\det(C' + i\eta J) = \det[S(C + i\eta J)S^T]$$
$$= \det(S)\det(S^T)\det(C + i\eta J) = \chi(\eta),$$

from which the invariance of the roots follows immediately. ■

Let the positive roots of $\chi(\eta) = 0$ be $\eta_1 \leqslant \eta_2 \leqslant \cdots \leqslant \eta_n$. We will call these roots the *Williamson invariants*. For the present, assume that these roots are all distinct. Because $\chi(\eta_j) = \det(C + i\eta_j J) = 0$, there will exist a vector $W_j \in \mathbb{C}^{2n}$ such that

$$(C + i\eta_j J)W_j = 0. \tag{4.4}$$

Taking the complex conjugate of both sides of (4.4) results in

$$(C - i\eta_j J)\bar{W}_j = 0. \tag{4.5}$$

Now, let

$$X_j = (W_j + \bar{W}_j)/2 \quad \text{and} \quad Y_j = (W_j - \bar{W}_j)/2i. \tag{4.6}$$

These two equations may be combined to give the following set of equations:

$$CX_j = \eta_j JY_j, \quad CY_j = -\eta_j JX_j. \tag{4.7}$$

Multiply both equations by X_k^T and then Y_k^T. Using the resulting set together with (1.1) yields

$$X_k^TCX_j = \eta_j\sigma(Y_j,X_k) \quad \text{and} \quad X_k^TCY_j = -\eta_j\sigma(X_j,X_k),$$
$$Y_k^TCX_j = \eta_j\sigma(Y_j,Y_k) \quad \text{and} \quad Y_k^TCY_j = -\eta_j\sigma(X_j,Y_k). \tag{4.8}$$

Comparing (4.8) with the set one gets by interchanging j, k in (4.8) and using the fact that $C^T = C$, we arrive at

$$\eta_j\sigma(Y_j,X_k) = \eta_k\sigma(Y_k,X_j)$$

and

$$\eta_j\sigma(Y_j,Y_k) = \eta_k\sigma(Y_k,Y_j)$$
$$\eta_j\sigma(X_j,X_k = \eta_k\sigma(X_k,X_j),$$

and

$$\eta_j\sigma(X_j,Y_k = \eta_k\sigma(X_k,Y_j). \tag{4.9}$$

Using (4.9), the antisymmetry of σ, and the assumption that the η_j's are distinct, we have, for $j \neq k$, that

$$\sigma(Y_j,X_k) = \sigma(X_j,X_k) = \sigma(Y_j,Y_k) = 0. \tag{4.10}$$

If $j = k$, then (4.8) implies that

$$\sigma(Y_j,X_j) = X_j^TCX_j/\eta_j = Y_j^TCY_j/\eta_j > 0. \tag{4.11}$$

Thus, by normalizing X_j so that $X_j^TCX_j = \eta_j$, we obtain

$$\sigma(Y_j,X_j) = 1. \tag{4.12}$$

Taken together, (4.10) and (4.12) imply that $\{X_1,...,X_n;Y_1,...,Y_n\}$ is a symplectic basis for $\mathbb{R}^{2n}$, and that the matrix

$$S \equiv (X_1 \ \cdots \ X_n \ \ Y_1 \ \cdots \ Y_n) \tag{4.13}$$

is symplectic [i.e., it satisfies (3.1)]. Moreover, a straightforward manipulation using (4.8), (4.10), (4.12), and (4.13) gives

$$S^TCS = \mathrm{diag}(\eta_1,\ldots,\eta_n \quad \eta_1,\ldots,\eta_n). \tag{4.14}$$

So far, we have assumed that the η_j's were distinct. What happens if we drop that assumption? Not much, really. Because both $C^{1/2}W_j$ and $C^{1/2}\overline{W}_j$ are eigenvectors of the Hermitian matrix $-iC^{-1/2}JC^{-1/2}$, the W_j's and $\overline{W}_j$'s compare a basis for $\mathbb{C}^{2n}$. The correpsonding real and imaginary parts, the X_j's and Y_j's, form a basis for $\mathbb{R}^{2n}$. When $\eta_j \neq \eta_k$, the arguments above still imply that (4.10) holds. When degeneracy occurs, a simple, obvious modification of the usual Gram–Schmidt process can be used to get (4.10) to hold for X's and Y's coming from linearly independent W_j and W_k corresponding to $\eta_j = \eta_k$. We leave the details to the reader. In any case, we have shown the following to be true.

Theorem 4.2: If C is a positive definite, real, symmetric $2n\times 2n$ matrix, then (4.14) holds with the symplectic matrix S constructed as above.

It should be pointed out that Theorem 4.2 could also be proved by starting with an orthonormal basis of eigenvectors for the Hermitian matrix $-iC^{-1/2}JC^{-1/2}$, and then using the orthonormality of this set plus the fact that the eigenvectors of this matrix are all of the form $C^{1/2}W$ or $C^{1/2}\overline{W}$ to deduce (4.10) and (4.12). We chose the method used here because it directly gets at the quantities of interest, the W's and $\overline{W}$'s.

We also wish to point out that even when an η_j has no degeneracy (that is, its eigenspace has one complex dimension), the phase of the corresponding W_j is arbitrary. A change $W_j \to e^{i\theta}W_j$ has the effect of the changes $X_j \to \cos\theta\, X_j - \sin\theta\, Y_j$ and $Y_j \to \sin\theta\, X_j + \cos\theta\, Y_j$. This freedom is due to rotations in $\mathbb{R}^2$ commuting with the 2×2 version of J, which in that case is itself a rotation through an angle $\pi/2$.

We conclude our matricial interlude with a result that will prove fundamental in what is to follow. It will provide us with a tool with which we can link the analytic and geometric characterizations of uncertainty.

Corollary 4.3: Let C be a real, symmetric positive definite $2n\times 2n$ matrix and let $\eta_1 \leqslant \eta_2 \cdots \leqslant \eta_n$ be the Williamson invariants for C. The matrix $C + i\eta J$ is non-negative if and only if

$$|\eta| \leqslant \eta_1. \tag{4.15}$$

Proof: From (3.11), (4.14), and the discussion at the end of Sec. III, we find that $C + i\eta J$ is non-negative if and only if the matrix

$$\mathrm{diag}(\eta_1,\ldots,\eta_n \quad \eta_1,\ldots,\eta_n) + i\eta J \tag{4.16}$$

is non-negative. It is a straightforward matter to check that the eigenvalues of (4.16) are all of the form

$$\lambda_j^{\pm} = \eta_j \pm \eta. \tag{4.17}$$

All of these eigenvalues will be non-negative, and hence (4.16) will be non-negative if and only if $|\eta| \leqslant \eta_j$, for $j = 1,\ldots,n$. Since η_1 is the smallest of the Williamson invariants, this condition is itself equivalent to (4.15). ■

V. INVARIANCE AND UNCERTAINTY

There are two things that we wish to do in this section. The first is to formulate the uncertainty principle in a way that is invariant under affine canonical transformations, and the second is to discuss the "directions" that give minimal uncertainty for a WDF.

Using the Williamson invariants discussed in the previous section, we can formulate an invariant uncertainty principle as follows.

Theorem 5.1: Let C be a real, symmetric, positive definite $2n\times 2n$ matrix. Then C is the covariance matrix for a WDF if and only if

$$\hbar/2 \leqslant \eta_1 \equiv \text{the smallest Williamson invariant of } C. \tag{5.1}$$

Proof: Apply Corollary 4.3, Theorem 2.4, and Proposition 4.1. ■

To get an idea of what this is saying, we will look at an example given by Littlejohn (see Ref. 1, § 2.2). Consider the Gaussian function

$$\rho_0(z) \overset{\text{def}}{=} (1/2\pi LK)e^{-(1/2)z^TC_0^{-1}z},$$

where

$$C_0 = \begin{pmatrix} L^2 & 0 \\ 0 & K^2 \end{pmatrix}. \tag{5.2}$$

In this case, we have C_0 as our candidate for a covariance matrix. The constants L and K are both positive. The matrix C_0 is clearly positive definite. To find its Williamson invariants, we need to get the positive roots of

$$\chi(\eta) = \det\begin{pmatrix} L^2 & i\eta \\ -i\eta & K^2 \end{pmatrix} = L^2K^2 - \eta^2 = 0. \tag{5.3}$$

There is only one positive root of (5.3), $\eta_1 = LK$, and this is the only Williamson invariant. Thus C_0 will be the covariance for a WDF if and only if $LK \geqslant \hbar/2$. Indeed, this condition and ρ_0's being a Gaussian are necessary and sufficient for ρ_0 to be a WDF.[11–14]

Since, as Littlejohn points out, $L = \Delta q$ and $K = \Delta p$ are the dispersions in the Gaussian ρ_0, the inequality (5.1) is simply the usual uncertainty relation in this case. Assume that $\eta_1 = LK \geqslant \hbar/2$ holds, so that ρ_0 is a WDF. Let us now look at what happens when we let ρ_0 evolve freely via the Hamiltonian $p^2/2m$. The WDF after time t is

$$\rho_t(z) = \rho_0(S_t^{-1}z), \quad \text{where } S_t \equiv \begin{pmatrix} 1 & t/m \\ 0 & 1 \end{pmatrix}. \tag{5.4}$$

Note that S_t is symplectic. From (3.12), the covariance matrix for ρ_t is

$$C_t = S_tCS_t^T = \begin{pmatrix} L^2 + K^2t^2/m^2 & K^2t/m \\ K^2t/m & K^2 \end{pmatrix}. \tag{5.5}$$

We can read off the dispersions in q and p for ρ_t from (5.5):

$$\Delta q_t = \sqrt{L^2 + K^2t^2/m^2} \quad \text{and} \quad \Delta p_t = K. \tag{5.6}$$

Obviously, we have that

$$\Delta q_t \Delta p_t \geqslant \eta_1 \geqslant \hbar/2. \tag{5.7}$$

This inequality suggests that η_1, which is the lowest Williamson invariant for C_t as well as for C_0 [because the two are related via (5.5)], plays the role of a minimal uncertainty in this example. In fact, we have a more general result.

Theorem 5.2: Let C be a symmetric, positive definite $2n\times 2n$ matrix with η_1 as its lowest Williamson invariant, and suppose that C is the covariance matrix for some WDF ρ. For every pair $x,y\in\Gamma$ such that $\sigma(y,x)=1$, the dispersions in the coordinates

$$Q\equiv\sigma(y,z) \quad\text{and}\quad P\equiv\sigma(z,x) \tag{5.8}$$

satisfy

$$\Delta Q\Delta P\geqslant\eta_1. \tag{5.9}$$

Finally, there exists a pair x and y for which equality holds in (5.9).

Proof: Let us first show that we can find a pair $x, y\in\Gamma$ for which equality holds in (5.9). Our covariance matrix C satisfies the conditions placed on C in Sec. IV. Let X_1 and Y_1 be defined by (4.6) with $j=1$. From (4.12)—again, with $j=1$, we have that $\sigma(Y_1,X_1)=1$. For x and y, we will take

$$x=JX_1 \quad\text{and}\quad y=JY_1. \tag{5.10}$$

It is clear that

$$\sigma(y,x)=(JX_1)^TJ(JY_1)=-X_1^TJ^3Y_1=\sigma(Y_1,X_1)=1.$$

With the help of (4.7) and using the fact that the expectation value of z with respect to ρ may be assumed to be 0, the square of the dispersion for Q can be calculated from the following sequence of steps:

$$\Delta Q^2=\int_\Gamma\sigma(y,z)^2\rho(z)dz=-y^TJ\left(\int_\Gamma zz^T\rho(x)dz\right)Jy=-(JY_1)^TJCJ^2Y_1=Y_1^TCY_1=-\eta_1Y_1^TJX_1=\eta_1\sigma(Y_1,X_1)=\eta_1. \tag{5.11}$$

A similar calculation shows that

$$\Delta P^2=\eta_1. \tag{5.11$'$}$$

Thus, for our choice of x, y, we see from (5.11) and (5.11′) that the uncertainty relation (5.9) holds with equality.

We will now show that (5.9) holds for arbitrary Q, P satisfying (5.8), as long as $\sigma(y,x)=1$. By means of matrix manipulations similar to the ones used above, we have that the matrix

$$C'\equiv\int_\Gamma\begin{pmatrix}Q^2 & QP\\ PQ & P^2\end{pmatrix}\rho(z)dz \tag{5.12}$$

can be written in the form

$$C'=A^TCA,\quad\text{where}\quad A=(Jy\quad -Jx). \tag{5.13}$$

Note that A is a $2n\times 2$ real matrix. Observe that since η_1 is the lowest Williamson invariant of C, the matrix $C=i\eta_1J$ is non-negative. From this it follows that the matrix $A^TCA+i\eta_1A^TJA$ is non-negative. We can put this matrix in a more useful form with the following sequence of steps:

$$\begin{aligned}A^TCA+i\eta_1A^TJA&=C'+i\eta_1\begin{pmatrix}-y^TJ\\ x^J\end{pmatrix}J(Jy\quad -Jx)\\ &=C'+i\eta_1\begin{pmatrix}\sigma(y,y) & -\sigma(x,y)\\ -\sigma(y,x) & \sigma(x,x)\end{pmatrix}\\ &=C'+i\eta_1\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}.\end{aligned} \tag{5.14}$$

Using the non-negativity of this last matrix, together with (5.13), the usual formulas for dispersions (again, assuming the expectation for z with respect to ρ is 0), and the standard determinant condition for a 2×2 matrix to be non-negative, we obtain

$$\Delta Q^2\Delta P^2\geqslant\eta_1^2+\left(\int_\Gamma QP\rho(z)dz\right)^2,$$

from which (5.9) follows immediately. ■

In the course of proving Theorem 5.2, we produced $x,y\in\Gamma$, with $\sigma(y,x)=1$, such that the corresponding conjugate coordinates Q, P defined by (5.8) have dispersions for which the uncertainty product is minimal. For future reference, we wish to write out x, y separately.

Corollary 5.3: With the notation and assumptions of Theorem 5.2, the $x,y\in\Gamma$ which give equality in (5.9) are

$$x=JX_1 \quad\text{and}\quad y=JY_1, \tag{5.15}$$

where X_1 and Y_1 are normalized so that $\sigma(Y_1,X_1)=1$ and satisfy

$$CX_1=\eta_1JY_1 \quad\text{and}\quad CY_1=-\eta_1JX_1. \tag{5.16}$$

Here, η_1 is the smallest Williamson invariant of C. This choice of x and y is, in general, not unique.

Proof: The nonuniqueness of x and y stems from the nonuniqueness of X_1 and Y_1. See the remarks following Theorem 4.2. ■

Let us return to Littlejohn's example. In particular, we want to look at the case in which $LK\geqslant\hbar/2$, so ρ_0 in (5.2) is a WDF. The WDF that evolves freely from ρ_0 is ρ_t; this is given in (5.4). The covariance matrix C_t for ρ_t is found in (5.5); its only Williamson invariant is $\eta_1=LK$. The best way to find X_1 and Y_1 for C_t is first to solve

$$(C_t+i\eta_1J)W_1=0. \tag{5.17}$$

This is easy to do and results in

$$W_1=c\begin{pmatrix}-iK\\ L+iKt/m\end{pmatrix}. \tag{5.18}$$

Here, c is a complex constant; its modulus will be determined by requiring $\sigma(Y_1,X_1)=1$, but its phase will remain arbitrary. We will therefore choose it so that $c>0$. With this choice of phase, we have

$$X_1 = \frac{W_1 + \overline{W}_1}{2} = \begin{pmatrix} 0 \\ cL \end{pmatrix}$$

and (5.19)

$$Y_1 = \frac{W_1 - \overline{W}_1}{2i} = \begin{pmatrix} -cK \\ cKt/m \end{pmatrix}.$$

Since $\sigma(Y_1, X_1) = c^2 LK$, we need to choose $c = 1/\sqrt{LK}$. Our final result for x and y is therefore

$$x = JX_1 = \sqrt{\frac{L}{K}}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad y = JY_1 = \sqrt{\frac{K}{L}}\begin{pmatrix} t/m \\ 1 \end{pmatrix}. \tag{5.20}$$

Of course, as we noted after Theorem 4.2, a different choice of phase will result in a rotation of the vectors X_1 and Y_1, and therefore of x and y as well.

Having gotten x and y, we can write down the coordinates Q and P from (5.8):

$$Q = \sigma(y,z) = \sqrt{K/L}\,(q - pt/m), \qquad P = \sigma(z,x) = \sqrt{L/K}\,p. \tag{5.21}$$

When $t = 0$, the Q-P frame is just a scaled version of the q-p frame. As t increases, the new frame moves to compensate exactly for the time evolution involved. A similar phenomenon would occur for any evolution generated by a quadratic Hamiltonian.

We close this section by pointing out that the directions of minimal uncertainty and the minimal uncertainty itself for a WDF are interesting quantities to look at. First of all, they enable one to separate what amounts to a classical effect, "spreading of the wave packet," from a true quantum mechanical change in the covariance. When a WDF is subjected to evolution stemming from a nonquadratic Hamiltonian, the lowest Williamson invariant η_1, which must be no smaller than $\hbar/2$, may either increase *or* decrease. To see that this must be so, one need only reflect on what happens when the nonquadratic Hamiltonian is chosen to be time-reversal invariant.

Second, knowing the directions of minimal uncertainty could be useful in that one could use a quadratic Hamiltonian to generate a linear canonical transformation that would take a WDF into one for which the quantities one wants to measure would have minimal uncertainty. This is the sort of thing one does with "squeezed" states.

Finally, one could take the idea outlined above a step further by working in the Williamson normal coordinates themselves. Something similar in spirit to doing that was done in Ref. 2 to get a convergent perturbation series for solutions to the quantum Liouville equation.

VI. GEOMETRY AND UNCERTAINTY

In Sec. V we dealt with an invariant version of the uncertainty principle and we discussed directions of minimal uncertainty. We will now turn to a geometric interpretation of the uncertainty relations.

Let C be a positive definite, real, symmetric $2n \times 2n$ matrix. If we suppose that its lowest Williamson invariant, η_1, satisfies $\eta_1 \geqslant \hbar/2$, then C will be the covariance of some WDF ρ; we may assume that the expectation of z relative to ρ is 0. Define the quadratic forms

$$w(z) \equiv \tfrac{1}{2} z^T C^{-1} z \quad \text{and} \quad w^*(a) \equiv \tfrac{1}{2} a^T C a. \tag{6.1}$$

The level surfaces of w are the *Wigner ellipsoids* corresponding to the covariance C. These ellipsoids are to be thought of as surrounding regions of Γ in which the WDF ρ is appreciable. The function w^* has the significance described below.

Proposition 6.1: The function w^* is the Legendre transform of w.

Proof: The Legendre transform (cf. Refs. 18 and 19) of w is given by

$$w^*(a) = a^T z - w(z), \tag{6.2}$$

where z is expressed in terms of a by solving the equation

$$a = \nabla w(z) = C^{-1} z, \tag{6.3}$$

for z in terms of a. The result is, of course, that

$$z = Ca. \tag{6.4}$$

Inserting (6.4) in (6.2) and simplifying, we get the expression for w^* given in (6.1). ■

Because of the relationship between w and w^*, we will call the level surfaces of w^*, which are themselves ellipsoids, *dual Wigner ellipsoids*. It is no accident that we have used the symbol a for the Legendre transform variable and for the symplectic Fourier transform variable. In both cases, a is to be thought of as belonging to Γ'. Dual Wigner ellipsoids may be regarded as surrounding regions of Γ' in which $\tilde{\rho}(a)$ is appreciable. (Recall that we have assumed that C is the covariance corresponding to the WDF ρ.)

When only one degree of freedom is present, there is a well-known, simple, geometric interpretation to the uncertainty relation: the area enclosed by the Wigner ellipse $w(z) = 1$ is greater than or equal to $\pi\hbar$. Indeed, when $n = 1$, this area is easily seen to be $\pi\sqrt{\det(2C)} = 2\pi\eta_1$, and so its being greater than or equal to $\pi\hbar$ is equivalent to the C's being a covariance for a WDF. For the area enclosed by the dual Wigner ellipse $w^*(a) = 1$, one can easily obtain a corresponding result: this area is *less than or equal to* $4\pi/\hbar$ if and only if C is a covariance for a WDF.

The geometric interpretation given above is important. Let S be a symplectic matrix. Recall that when we transform $\rho(z) \to \rho(S^{-1}z) = \rho'(z)$, the covariance transforms this way: $C \to SCS^T = C'$. The Wigner ellipsoid for C' is obviously the one for C subjected to the linear canonical transformation $z \to Sz$. Similarly, the dual Wigner ellipsoid for C' is the dual Wigner ellipsoid for C subjected to the linear canonical transformation $a \to S^T a$. For $n = 1$, areas are invariant under such transformations. Thus, for the case of one degree of freedom, the geometic form of the uncertainty relation is invariant under linear canonical transformations.

We want to generalize this result to *all n*. Working with areas *per se* will not do. Areas remain invariant under symplectic transformations only when $n = 1$. To get the correct generalization, we point out that both of the areas mentioned above are directly related to the *first Poincaré invariant* for the ellipses involved.

If γ is a closed, piecewise smooth curve in Γ, and if γ forms the boundary of a piecewise smooth, orientable two-dimensional surface Σ in Γ, then

$$\mathscr{I}_1(\gamma) \equiv \frac{1}{2}\int_\gamma \sigma(dz,z) = \int_\Sigma \sum_{j=1}^n dq_j \wedge dp_j \tag{6.5}$$

is the first Poincaré invariant of γ.[18,20] We can easily compute this invariant for the ellipses resulting from the intersection of a subspace and either the Wigner ellipsoid or the dual Wigner ellipsoid.

Proposition 6.2: Consider the subspaces $\Pi\subset\Gamma$ and $\Pi^*\subset\Gamma'$ with bases $B=\{x,y\}$ and $B^*=\{X,Y\}$. If γ and γ^* are oriented ellipses with traces

$$\text{trace }\gamma = \Pi\cap\{z\in\Gamma|w(z)=1\}$$

and

$$\text{trace }\gamma^* \equiv \Pi\cap\{a\in\Gamma'|w^*(a)=1\}, \tag{6.6}$$

then we have

$$\mathscr{I}_1(\gamma) = \frac{2\pi\sigma(y,x)}{\sqrt{\det(A^TC^{-1}A)}}$$

and (6.7)

$$\mathscr{I}_1(\gamma^*) = \frac{2\pi\sigma(Y,X)}{\sqrt{\det(A^{*T}CA^*)}},$$

where A and A^* are the column matrices $A=(x\quad y)$ and $A^*=(X\quad Y)$.

Proof: Both expressions are obtained in nearly identical fashion. We will derive only the one for $\mathscr{I}_1(\gamma)$. Since B is a basis for Π, every $z\in\Pi$ can be written as a linear combination of x and y. Thus there are constants r, s such that

$$z = rx + sy. \tag{6.8}$$

If z also is in γ, the intersection of Π and the Wigner ellipse $w(z)=1$, then the point $(r,s)\in\mathbb{R}^2$ is on the ellipse

$$(r\quad s)A^TC^{-1}A\begin{pmatrix}r\\ s\end{pmatrix} = 2. \tag{6.9}$$

We will let γ_B be the ellipse (6.9) oriented counterclockwise relative to the r-s coordinates. Substituting (6.8) in (6.5) and using Green's theorem, we get

$$\mathscr{I}_1(\gamma) = \frac{\sigma(y,x)}{2}\int_{\gamma_B} r\,ds - s\,dr = \sigma(y,x)\,[\text{area in }\gamma_B]. \tag{6.10}$$

On the other hand, from (6.9) we see that

$$\text{area in }\gamma_B = 2\pi/\sqrt{\det(A^TC^{-1}A)}. \tag{6.11}$$

Putting all this together yields the formula for $\mathscr{I}_1(\gamma)$ in (6.7). ■

Taking the subspace Π_j spanned by $B_j=\{x_j,y_j\}$, where $x_j=JX_j$ and $y_j=JY_j$, with X_j and Y_j being defined in (4.6), let γ_j be the oriented ellipse coming from the intersection of Π_j and the Wigner ellipsoid $w(z)=1$. From (4.7), (4.10), and (4.11), we find that with

$$A_j \equiv (JX_j\quad JY_j), \tag{6.12}$$

we have that

$$A_j^TC^{-1}A_j = \frac{1}{\eta_j}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}. \tag{6.13}$$

Finally, using (6.13) in (6.7) yields

$$\mathscr{I}_1(\gamma_j) = 2\pi\eta_j. \tag{6.14}$$

In this calculation, we never assumed that C was a covariance for a WDF. We only assumed that C was positive definite. By combining (6.14) with Theorem 5.1, we therefore arrive at the following.

Theorem 6.3: Suppose that C is a real, symmetric, positive definite $2n\times 2n$ matrix, that $\{X_1,\ldots,X_n,Y_1,\ldots,Y_n\}$ is the Williamson basis corresponding to C, and that $\Pi_j\subseteq\Gamma$ is the subspace with basis $\{x_j=JX_j, y_j=JY_j\}$, where $j=1,\ldots,n$. With some orientation, each ellipse $\gamma_j=\Pi_j\cap\{z\in\Gamma|w(z)=1\}$ satisfies

$$\mathscr{I}_1(\gamma_j)\geqslant\pi\hbar,\quad\text{where } j=1,\ldots,n, \tag{6.15}$$

if and only if C is a covariance matrix for some WDF.

This result generalizes that for the Wigner ellipse in the $n=1$ case. It is, however, disappointing. One would like to have a result that holds for the intersection of an arbitrary subspace and a Wigner ellipsoid. It is not hard to show that, in Γ, there are many subspaces for which (6.15) simply fails. For example, if Π is a *null* (*isotropic*) subspace, then, no matter what basis is chosen, one always has $\sigma(y,x)=0$, and so $\mathscr{I}_1(\gamma)=0$ in that case. As the following theorem shows, the situation for the dual Wigner ellipsoid is much better.

Theorem 6.4: Let C be a positive definite $2n\times 2n$ symmetric matrix and let $w^*(a)$ be as in (6.1). Then C is the covariance of a WDF if and only if for every two-dimensional subspace $\Pi^*\subseteq\Gamma'$ the first Poincaré invariant for the ellipse

$$\gamma^*\equiv\Pi^*\cap\{a\in\Gamma'|w^*(a)=1\}$$

satisfies

$$\mathscr{I}_1(\gamma^*)\leqslant 4\pi/\hbar, \tag{6.16}$$

provided the orientation of γ^* is chosen correctly.

Proof: Let $B^*\equiv\{X,Y\}$ be a basis for Π^*. We assume that $\sigma(Y,X)$ is non-negative. (This can be arranged by relabeling X and Y.) A vector a will be in Π^* if and only it has the form

$$a = uX + vY,\quad u,v\in\mathbb{R}. \tag{6.17}$$

Let A^* be the $2n\times 2$ matrix

$$A^* = (X\quad Y). \tag{6.18}$$

It is clear that $a\in\gamma^*$ if and only if u and v satisfy

$$(u\quad v)A^{*T}CA^*\begin{pmatrix}u\\ v\end{pmatrix} = 2, \tag{6.19}$$

which is itself an ellipse in u-v space. We will denote this ellipse, when traversed in the positive direction relative to u-v coordinates, by $\gamma^*_{B^*}$. (Assigning an orientation to $\gamma^*_{B^*}$ automatically assigns one to γ^*.)

If η_1 is the smallest Williamson invariant of C, then, by an argument similar to that used in proving Theorem 5.2, we have that

$$A^{*T}CA^* + i\eta_1\sigma(Y,X)\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix} \tag{6.20}$$

is non-negative. By Corollary 4.3, we see that

$$\sigma(Y,X)\eta_1 \leqslant \text{the smallest Williamson invariant of } A^{*T}CA^{*}. \tag{6.21}$$

Since $A^{*T}CA^{*}$ is a 2×2 matrix, the right side of (6.21) is just $\sqrt{\det(A^{*T}CA^{*})}$, and so, after rearranging terms and multiplying by 2π, we obtain

$$2\pi\sigma(Y,X)/\sqrt{\det(A^{*T}CA^{*})} \leqslant 2\pi/\eta_1. \tag{6.22}$$

By Proposition 6.2, the left side of (6.22) is $\mathscr{J}_1(\gamma^*)$. Thus we have

$$\mathscr{J}_1(\gamma^*) \leqslant 2\pi/\eta_1. \tag{6.23}$$

Consider the special case of the subspace $\Pi^* = \Pi_1^*$ with $X = X_1$ and $Y = Y_1$, where X_1 and Y_1 are as in (4.6). With this choice of Π^*, a computation nearly identical to that used in proving Theorem 6.3 yields

$$\mathscr{J}_1(\gamma^*) = 2\pi/\eta_1, \quad \text{when } \Pi^* = \Pi_1^*. \tag{6.24}$$

We can now complete our proof. First of all, if (6.16) is satisfied for all possible Π^*, then it is satisfied with $\Pi^* = \Pi_1^*$. Putting (6.16), with $\Pi^* = \Pi_1^*$, together with (6.24), we see that $\eta_1 \geqslant \hbar/2$. Theorem 5.1 then implies that C is the covariance for some WDF.

Conversely, if C *is* a covariance for a WDF, then again by Theorem 5.1, we have $\eta_1 \geqslant \hbar/2$. Thus (6.23) implies the inequality in (6.16). ■

As we remarked earlier, Theorem 6.3 is not as nice a characterization of the Wigner ellipsoid as Theorem 6.4 is of the dual Wigner ellipsoid. At first, this may seem surprising, but in fact it is what one should expect. The covariance matrix contains *nonlocal* information about the WDF ρ that gives rise to it, whereas it contains *local* information about $\tilde{\rho}$. Cutting either ellipsoid with a plane through the origin in Γ and looking at the first Poincaré invariant for the resulting ellipse is an operation that is local in character, and clearly works better when the information sought after is itself local.

All that we have said above has focused on the *first* Poincaré invariant. Are there any results for the *higher*-order invariants? For the Wigner ellipsoid itself, a few things are known. For example, the volume of the Wigner ellipsoid, $\mathscr{J}_n$, must be no smaller than $(\pi\hbar)^n$. But apart from results similar in character to Theorem 6.3, not much can be said. As one might suspect from our earlier discussion, the situation with respect to the *dual* Wigner ellipsoid is better.

Consider a null subspace $\mathscr{N}\subset\Gamma'$. We know that the dimension of $\mathscr{N}$, which we denote by k, cannot exceed n (see Ref. 18, p. 223). Suppose that $\{E_1,\ldots,E_k\}$ is a basis for $\mathscr{N}$. It is easy to show that there exists a second null subspace $\mathscr{M}$ (which is not unique) with these properties: (i) $\mathscr{M}\cap\mathscr{N} = \varnothing$; and (ii) $\mathscr{M}$ has a basis $\{F_1,\ldots,F_k\}$ for which $\sigma(F_j,E_{j'}) = \delta_{jj'}$. Let $\Pi^* = \mathscr{N} + \mathscr{M} \subseteq \Gamma'$ be the subspace with basis $B^* = \{E_1,\ldots,E_n; F_1,\ldots,F_n\}$. We then have this, our final result.

Theorem 6.5: Let γ^* be the ellipsoid $\Pi^*\cap\{a\in\Gamma' \mid w^*(a) = 1\}$ together with some orientation. In addition, suppose that C is a real, symmetric, positive definite $2n\times 2n$ matrix. If η_1 is C's lowest Williamson invariant then

$$\mathscr{J}_k(\gamma^*) \leqslant (2\pi/\eta_1)^k. \tag{6.25}$$

If, in addition, C is a covariance for a WDF, then

$$\mathscr{J}_k(\gamma^*) \leqslant (4\pi/\hbar)^k. \tag{6.26}$$

Proof: Let Σ be the interior of γ^* in Π. By Stokes' theorem, the kth Poincaré invariant is

$$\mathscr{J}_k(\gamma^*) = \int_\Sigma \underbrace{\omega\wedge\cdots\wedge\omega}_{k\text{ times}}, \tag{6.27}$$

where $\omega = d\left[\tfrac{1}{2}\sigma(da,a)\right]$ is the invariant two-form on Γ'. Since $a\in\Pi^*$, we can write as a linear combination of the basis vectors; that is,

$$a = \sum_{j=1}^{k} r_j E_j + s_j F_j.$$

In these coordinates,

$$\omega = \sum_{j=1}^{k} ds_j \wedge dr_j,$$

and so, using the properties of the basis vectors, we get

$$\underbrace{\omega\wedge\cdots\wedge\omega}_{k\text{ times}} = k!\prod_{j=1}^{k} ds_j\wedge dr_j. \tag{6.28}$$

Combining (6.27) and (6.28) results in

$$\mathscr{J}_k(\gamma^*) = k!\,[\text{volume of } \Sigma_{B^*}], \tag{6.29}$$

where Σ_{B^*} is the interior of the ellipsoid

$$(r_1 \ \cdots \ s_k)A^{*T}CA^{*}\begin{pmatrix} r_1 \\ \vdots \\ s_k \end{pmatrix} = 2,$$

where

$$A = (E_1 \ \cdots \ E_k \ \ F_1 \ \cdots \ F_k). \tag{6.30}$$

One may do a standard calculation to get the volume of Σ_{B^*}. Performing this calculation and using (6.29) and (6.30), we arrive at

$$\mathscr{J}_k(\gamma^*) = (2\eta)^k/\sqrt{\det A^{*T}CA^{*}}. \tag{6.31}$$

Using the properties of the basis B^* and employing an argument similar to that used in Theorem 5.2, we can easily show that the matrix $A^{*T}CA^{*} + i\eta_1 J_k$ is non-negative. [Here, J_k is the J in (1.1′) with $n\to k$.] Corollary 4.3 then implies that η_1 is no larger than the smallest Williamson invariant of $A^{*T}CA^{*}$. From the theory constructed in Sec. IV, it is easy to show that

$$\sqrt{\det A^{*T}CA^{*}} = \text{product Williamson invariants of}$$

$$A^{*T}CA^{*} \geqslant \eta_1^k \tag{6.32}$$

Combining (6.31) and (6.32) yields (6.25). If C is the covariance for a WDF, then (6.25) and Theorem 5.1 imply (6.26). ■

[1] R. G. Littlejohn, "The semiclassical evolution of wave packets," Phys. Rep. **138**, 193 (1986).

[2] F. J. Narcowich, " A Dyson-like expansion for solutions to the quantum Liouville equation," J. Math. Phys. **27**, 2502 (1986).

[3] F. J. Narcowich and R. F. O'Connell, "Necessary and sufficient conditions for a function defined on phase space to be a Wigner distribution," Phys. Rev. A **34**, 1 (1986).

[4]M. Hillery, R. F. O'Connell, M. O. Scully, and E. P. Wigner, "Distribution functions in physics: Fundamentals," Phys. Rep. **106**, 121 (1984).
[5]D. Kastler, "The C^*-algebras of a free boson field," Commun. Math. Phys. **1**, 14 (1965).
[6]G. Loupias and S. Miracle-Sole, "C^*-algébres des systéms canoniques I," Commun. Math. Phys. **2**, 31 (1966).
[7]G. Loupias and S. Miracle-Sole, "C^*-algébres des systéms canoniques II," Ann. Inst. H. Poincaré **6**, 39 (1967).
[8]M. D. Srinivas and E. Wolf, "Some nonclassical features of quantum mechanics," Phys. Rev. D **11**, 1477 (1975).
[9]F. J. Narcowich, "The problem of moments in the phase-space formulation of quantum mechanics," J. Math. Phys. **28**, 2873 (1987).
[10]E. Breitenberger, "Uncertainty measures and uncertainty relations for angle observables," Found. Phys. **15**, 353 (1985).
[11]F. J. Narcowich, "Conditions for the convolution of a two Wigner distributions to be itself a Wigner distribution," J. Math. Phys, **29**, 2036 (1988).
[12]F. J. Narcowich and R. F. O'Connell, "A unified approach to quantum dynamical maps and Gaussian Wigner distributions," Phys. Lett. A **133**, 167 (1988).
[13]H. P. Yuen, "Multimode two-photon coherent states," in *International Symposium on Spacetime Symmetries*, edited by Y. S. Kim and W. W. Zachary (North-Holland, Amsterdam, 1989).
[14]R. Simon, E. C. G. Sudarshan, and N. Mukunda, "Gaussian Wigner distributions: A complete characterization," Phys. Lett. A **124**, 223 (1987); "Gaussian Wigner distributions in quantum mechanics and optics," Phys. Rev. A **36**, 3868 (1987).
[15]G. Lindblad, "Gaussian quantum stochastic processes on the CCR algebra," J. Math. Phys. **20**, 2081 (1979).
[16]F. J. Narcowich, "A quantum mechanical moment problem," in *The Physics of Phase Space, Lecture Notes in Physics*, Vol. 278, edited by Y. S Kim and W. W. Zachary (Springer, Berlin, 1987).
[17]J. Williamson, "On the algebraic problem concerning the normal forms of linear dynamical systems," Am. J. Math. **58**, 141 (1936).
[18]V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978).
[19]R. T. Rockafellar, *Convex Analysis* (Princeton U. P., Princeton, NJ, 1972).
[20]H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, MA, 1950).

On nonperfect fluid cosmologies evolving towards perfect fluid cosmological models

J. Carot[a)]

Department of Mathematical Sciences, University of Aberdeen, The Edward Wright Building, Dunbar Street, Aberdeen AB9 2TY, Scotland, United Kingdom

(Received 15 June 1989; accepted for publication 30 August 1989)

The problem of entropy production in near-equilibrium situations and their evolution towards a state of thermodynamic equilibrium is considered in the cosmological context. A physically realistic model (i.e., satisfying the energy conditions) describing such a situation is constructed. From a few hypotheses and considerations, it is seen that the metric tensors of the space-time at both equilibrium and nonequilibrium configurations are conformally related. The material content described by the energy-momentum tensor is interpreted as a viscous fluid smoothly evolving into a perfect one.

I. INTRODUCTION

Cosmological models are intended to provide suitable representations of the universe in terms of the geometry of the space-time and those measurable magnitudes that are of physical interest (energy density, pressure, etc.).

A first approach to the problem (both historically and in terms of simplicity) is to consider the content of the universe as an isentropic perfect fluid obeying an equation of state of the form $p = p(\rho)$, p being the pressure and ρ the energy density. In general, this seems to be a good approximation, at large scale, to the present state of the universe (for general reference see for instance Ref. 1); however, this approach breaks down when it comes to describe the universe at earlier stages, when entropy production phenomena occurred and therefore thermodynamic equilibrium was not possible (unlike the situation described by an isentropic perfect fluid). Furthermore, many well-known hydrodynamic processes such as turbulence, cavitation, and shockwaves cannot take place in a perfect-fluid-filled universe.[2] An obvious next step in order to account for dissipative phenomena is to replace the perfect fluid with a more general type of matter allowing production of entropy; in particular, nonperfect-fluid models have been used, the production of entropy being then caused by viscous heating and/or heat transport. (For the effect of viscosity on cosmological models see for instance Refs. 3–6, and for the effect of the heat flux see Refs. 7–13.) These models provide good descriptions of situations not far from thermodynamic equilibrium (since only first-order deviations from the state of equilibrium are considered[14]), but definitely out of it.

The main purpose of this paper is to find a cosmological solution that evolves continuously from a nonequilibrium situation into an equilibrium, perfect-fluid situation, insisting that the transition between both states be smooth.

The paper is organized as follows: In Sec. II the general formalism is presented and developed, showing that, after a few hypotheses and considerations about the Weyl and energy-momentum tensors describing the above situations, one can conclude that the metric tensors of the space-time in both situations are conformally related, and that the material content of the space-time at the initial stage can be assimilated to a viscous fluid without heat conduction. The condition about smoothness in the transition between both situations results then in a set of conditions on the conformal factor and its derivatives. It is shown that the viscosity plays a significative role in the evolution of the model both physically and geometrically, since it turns out that the conformal factor can be expressed as a function of the viscosity alone. The relationship between the conformal factor and the entropy production density is also briefly discussed.

In Sec. III a brief review is made of some concepts in elementary thermodynamics, and they are specified according to the particular situation outlined in Sec. II. The energy conditions[15,16] are also dealt with in a similar way by using the results given in Ref. 17. This section contains virtually no new results (apart from the specifications corresponding to our particular case), but is included here for the sake of completeness.

Section IV contains a brief discussion on the possible isometry groups admitted by such a model. Finally, Sec. V presents an example of spatially homogeneous Bianchi type III viscous-fluid cosmology that smoothly degenerates into one of the perfect-fluid cosmological models (also of Bianchi type III) given in Ref. 18, and it is shown that in this case the conformal factor can be expressed as a function of the entropy production density.

II. GENERAL CONSIDERATIONS

As it was already pointed out, the aim of this paper is to obtain a cosmological solution accounting for a rather general situation where irreversible processes (i.e., positive entropy production) can take place, and that evolves towards an isentropic (constant entropy) state whose material content can be described, as usual, by an (isentropic) perfect fluid. The metric tensors describing both situations (the initial one characterized by entropy production and the final one of constant entropy) will be, in principle, different, since

[a)] On leave from Departament de Física, Universitat Illes Baleares, 07071 Palma de Mallorca, Spain.

their energy-momentum tensors are different.

On the other hand, we expect the matter content to be physically reasonable (i.e., to verify the energy conditions[15,16]); this restricts the possible Segrè types for the energy-momentum tensor describing the material content of the space-time at the initial, nonisentropic stage, to either $\{1,111\}$ or $\{2,11\}$ (or any of their degeneracies[19]), whereas the energy-momentum tensor of the perfect fluid describing the final situation is of the Segrè type $\{1,(111)\}$. However, it is easy to see by analyzing the canonical forms for the above Segrè types[19] that a tensor of Segrè type $\{2,11\}$ cannot change continuously into a tensor of Segrè type $\{1,(111)\}$; therefore the energy-momentum tensor of the space-time at the first stage must be of the Segrè type $\{1,111\}$ (or some degeneracy). This can be interpreted as corresponding to a general (imperfect) fluid (with nonisotropic pressures and/or heat conduction).

If we succeed in smoothly matching the metric tensors corresponding to the two different stages, we can interpret the situation as that of a fluid, imperfect at some initial stage, which evolves into a perfect-fluid state. We shall assume the world lines of the particles of the fluid to be smooth, future-directed timelike curves; therefore an observer co-moving with the fluid would describe the material content of the space-time by means of the following energy-momentum tensors[20]:

$$T_{ab} = (\rho + p)u_a u_b + p g_{ab} + \pi_{ab} + q_a u_b + u_a q_b \quad (1)$$

for the inital, nonperfect fluid stage; and

$$\underset{0}{T}_{ab} = (\rho_0 + p_0)\underset{0}{u}_a \underset{0}{u}_b + p_0 \underset{0}{g}_{ab} \quad (2)$$

for the fluid at its final, perfect stage.

The quantities ρ, p, ρ_0, and p_0 stand for the energy density and isotropic pressure, respectively, in both nonperfect- and perfect-fluid cases; u_a and $\underset{0}{u}_a$ designate the velocity field of the fluid in both cases; q_a and π_{ab} designate, respectively, the heat conduction vector and the anisotropic pressure tensor (satisfying $g^{ab}\pi_{ab} = \pi_{ab}u^b = q_a u^a = 0$); and g_{ab} and $\underset{0}{g}_{ab}$ are the metric tensors in both cases.

We shall assume the material content of the space-time known, in the sense of considering both energy-momentum tensors completely specified; therefore the metric tensors will be determined (up to a constant scaling factor) once their respective Weyl tensors are given.[21] We shall take both Weyl tensors to be equal (as functions of the coordinates). Roughly speaking, the Weyl tensor accounts for the part of the curvature not due to the material content, but to the vacuum. Choosing both Weyl tensors equal means, therefore, that the part of the curvature of the space-time due to the vacuum does not change as the matter evolves, or, in other words, changes in the curvature of the space-time are the consequence of changes (evolution) in the material content.

The choice we made implies that unless the Weyl tensor is Petrov type N,[22] the metric tensors g_{ab} and $\underset{0}{g}_{ab}$ must be conformally related. We can then write

$$g_{ab} = e^{2\phi}\underset{0}{g}_{ab} \equiv \Omega^2 \underset{0}{g}_{ab}, \quad (3)$$

where $\phi \in \Lambda^0(M)$ is some function of the coordinates.

The isentropic perfect-fluid stage with metric $\underset{0}{g}_{ab}$ will be achieved when $\phi = \text{const} = 0$ $(\Omega = 1)$, while the initial, nonisentropic stage will correspond to $\phi \neq$ const. Thus one can expect in principle some correspondence between the entropy and the conformal factor.

The velocity fields of the fluid at both stages then satisfy[23]

$$u^a = e^{-\phi}\underset{0}{u}^a, \quad (4)$$

and therefore the kinematical magnitudes that characterize the fluid—shear, rotation, expansion and acceleration—are related by[23]

$$\sigma_{ab} = e^{\phi}\underset{0}{\sigma}_{ab}, \quad (5a)$$

$$w_{ab} = e^{\phi}\underset{0}{w}_{ab}, \quad (5b)$$

$$\theta = e^{-\phi}\{\theta_0 + 3\underset{0}{g}^{ab}\underset{0}{u}_a\phi_b\}, \quad (5c)$$

$$\dot{u}_a = \underset{0}{\dot{u}}_a + \underset{0}{u}_a(\underset{0}{g}^{ab}\underset{0}{u}_a\phi_b) + \phi_a. \quad (5d)$$

We shall assume the fluid to move along geodesics, i.e., $\dot{u}_a = \underset{0}{\dot{u}}_a = 0$. This restriction implies that the gradient ϕ_a of the conformal factor must be parallel to the velocity field of the fluid, and therefore the fluid is irrotational: $w_{ab} = \underset{0}{w}_{ab} = 0$. We shall then write

$$\phi_{,a} = \mu \underset{0}{u}_a. \quad (6)$$

In particular, one may choose the time coordinate adapted to the four-velocity of the fluid,

$$\underset{0}{u}^a = \delta^a_t. \quad (7)$$

Since $\underset{0}{u}^a$ is geodesic and hypersurface orthogonal, one has

$$\underset{0}{ds}^2 = -dt^2 + \underset{0}{h}_{\alpha\beta}(t,x^\gamma)dx^\alpha dx^\beta, \quad (8)$$

and therefore, from (6) and (7),

$$\underset{0}{u}_a = -\delta^t_a, \quad (9)$$

$$\mu = -\phi_t. \quad (10)$$

We see immediately that in the above coordinate system the function ϕ depends only on time, and so does μ. (As a matter of fact, μ can be expressed as a function of ϕ; $\mu \equiv \mu(\phi)$, and ϕ itself can be used as time coordinate.)

One can now easily derive the explicit expression of the energy-momentum tensor T_{ab} given in (1) in terms of $\underset{0}{T}_{ab}$ and of ϕ and its derivatives simply by taking into account the relationship between the Ricci tensors of the two related metrics g_{ab} and $\underset{0}{g}_{ab}$ (see, for example, Refs. 16 and 23). One has

$$T_{ab} = \underset{0}{T}_{ab} - 2\phi_{tt}\underset{0}{u}_a\underset{0}{u}_b + 2\phi_t\underset{0}{u}_{a/b} + 2\phi_t^2\underset{0}{h}_{ab}$$

$$- \{2\phi_{tt} + 3\phi_t^2 + 2\phi_t\theta_0\}\underset{0}{g}_{ab}, \tag{11}$$

where $\underset{0}{u}_{a/b} \equiv \overset{0}{\nabla}_b \underset{0}{u}_a$ stands for the covariant derivative of $\underset{0}{u}_a$ with respect to the metric $\underset{0}{g}_{ab} \equiv \underset{0}{g}_{ab}$, and $\underset{0}{h}_{ab} + \underset{0}{u}_a\underset{0}{u}_b$ is the orthogonal projector to the velocity field.

The energy density ρ, isotropic pressure p, heat conduction q^a, and anisotropic pressure tensor π_{ab} appearing in (1) can now be easily evaluated from (11) (see for instance Ref. 24), and one has

$$\rho = e^{-2\phi}\{\rho_0 + 3\phi_t^2 + 2\phi_t\theta_0\}, \tag{12}$$

$$p = e^{-2\phi}\{p_0 - 2\phi_{tt} - \phi_t^2 - \tfrac{4}{3}\phi_t\theta_0\}, \tag{13}$$

$$q^a = 0, \tag{14}$$

$$\pi_{ab} = -2\mu e^{-\phi}\sigma_{ab}. \tag{15}$$

From (15) it is immediate to see that π_{ab} corresponds to the anisotropic pressure tensor of a viscous fluid and, therefore, that $\mu e^{-\phi}$ may be identified with the coefficient of kinematic viscosity η:

$$\eta \equiv \mu \cdot e^{-\phi}. \tag{16}$$

From (11) and (3) one can see that in addition to $\phi(t_1) = 0$, the conditions $\phi_{tt}(t_1) = \phi_t(t_1) = 0$ are required in order to match g_{ab} and $\underset{0}{g}_{ab}$, and T_{ab} and $\underset{0}{T}_{ab}$ continuously at some hypersurface $t = t_1$.

Since μ can be expressed as a function of ϕ, and ϕ is a function of t alone, the inverse function theorem when applied to (16) allows us to express the conformal factor as a function of the viscosity:

$$\Omega = f(\eta) = \begin{cases} f(\eta), & \eta > 0, \\ 1, & \eta \leqslant 0, \end{cases} \tag{17}$$

and consequently

$$ds^2 = f^2(\eta)\, \underset{0}{ds^2}. \tag{18}$$

We can thus regard (18) as the metric of a space-time filled with a viscous fluid that changes into a perfect fluid as the viscosity dies out.

As a final remark to this section, it is worth noticing that the Raychaudhuri equation[16,23] is identically satisfied for the viscous fluid provided it is satisfied for the perfect fluid, and vice versa.

III. ENTROPY AND ENERGY CONDITIONS

In this section we shall briefly review some concepts in relativistic thermodynamics within the framework of the Eckart[25] (or Landau and Lifshitz[26]) theory for relativistic imperfect fluids. (For further details see for instance Ref. 27, and for limitations and improvements to this model see Refs. 28–31.) In order to define the relevant thermodynamical magnitudes we shall limit ourselves to a portion of the fluid small enough to be considered as in thermodynamic equilibrium (although interacting with the rest of the system, so that the whole system will not, in general, be in equilibrium). We shall consider all the intensive thermodynamical magnitudes defined in the local Minkowskian frame co-moving with the fluid element under consideration.

In any thermodynamic system there are always some constituents conserved through any transformation the system undergoes; in this case those constituents are the baryons and their rest masses. This suggests study of the fluid element as a microcanonical ensemble, the associated thermodynamical potential being then the entropy.[32] The entropy s (per unit of baryon mass) will then be a function of the internal energy ϵ (per unit of baryon mass) and the specific volume v ($v \equiv n^{-1}$, n being the baryon mass density):

$$s = s(\epsilon, v). \tag{19}$$

All the other thermodynamical magnitudes can be obtained from $s(\epsilon, v)$ as[32]

$$\partial s/\partial \epsilon = 1/T, \tag{20}$$

$$\partial s/\partial v = \tilde{p}/T, \tag{21}$$

where T is the temperature and $\tilde{p}$ is the thermodynamic pressure. The latter is related to the isotropic pressure p occurring in (1) [and in our case given by (13)] through the equation[27]

$$p = \tilde{p} - \zeta\theta, \tag{22}$$

where ζ is the so-called bulk viscosity coefficient and θ is the fluid expansion. On the other hand, the total energy density ρ appearing in (1) is[32]

$$\rho = n(1 + \epsilon). \tag{23}$$

The interaction of the fluid element with the rest of the system can be described by the following three differential laws:

(i) Conservation of the baryon number (generalized mass conservation law),

$$(nu^a)_{;a} = 0; \tag{24}$$

(ii) Conservation of the energy and momentum,

$$T^{ab}{}_{;b} = 0, \tag{25}$$

which can be seen as the generalization of the first law of thermodynamics (note that in our case $\overset{0}{\nabla}_b \underset{0}{T}^{ab} = 0$ if and only if $\nabla_b T^{ab} = 0$ as an immediate consequence of $\underset{0}{g}_{ab}$ and g_{ab} being conformally related); and

(iii) The entropy production law, which constitutes a generalization of the second law of thermodynamics—

$$\hat{s} \equiv s^a{}_{;a} \geqslant 0, \tag{26}$$

where the entropy current s^a is defined as[27]

$$s^a \equiv nsu^a + T^{-1}/q^a. \tag{27}$$

From (27) and (1), a detailed evaluation of s can be made, giving as a result

$$\hat{s} \equiv T^{-1}\{\zeta\theta^2 - \pi_{ab}\sigma^{ab} - q^a(T^{-1}T_{,a} + \dot{u}_a)\}. \tag{28}$$

This equation shows up the dissipative character of the terms π_{ab} and $u_a q_b + q_a u_b$ in the energy-momentum tensor.

In the present case and at the nonperfect fluid stage (when entropy production occurs) we have $\pi_{ab} = -2\eta\sigma_{ab}$ and $q_a = 0$. Furthermore, we shall assume $\zeta = 0$. Such an assumption finds its justification in that this is so when the fluid we are dealing with is such that the trace T of its energy-momentum tensor can be expressed as a function of ϵ and/or n.[27] As this happens to be true for the fluid at its perfect stage, we assume, without much ado, that the same holds true throughout the previous nonperfect stage. Equation (28) then takes the form

$$T\hat{s} \equiv 2\eta\sigma_{ab}\sigma^{ab} \equiv \eta\sigma^2 , \quad (29)$$

and therefore states that the entropy is produced by viscous heating alone, dropping to zero when the viscosity does or, equivalently, it takes the form of Eq. (17) when the conformal factor takes the value 1.

Let us next analyze the energy conditions[15,16] that the fluid under consideration has to satisfy in order to be physically realistic. We shall follow the results obtained by Kolassis *et al.*[17] for the energy conditions in the special case of a viscous fluid with or without heat conduction. In the case of $q^a = 0$, a set of sufficient conditions for the weak and dominant energy conditions to be satisfied consists of[17]

$$\rho + p \geqslant (2/\sqrt{3})\eta\,|\sigma| , \quad (30)$$

$$\rho - p \geqslant (2/\sqrt{3})\eta\,|\sigma| , \quad (31)$$

when $\gamma \equiv \frac{8}{3}\sigma^a_b\sigma^b_c\sigma^c_a \neq 0$. For $\gamma = 0$ the necessary and sufficient conditions are

$$\rho + p \geqslant \eta\,|\sigma| , \quad (32)$$

$$\rho - p \geqslant \eta\,|\sigma| . \quad (33)$$

These conditions (30)–(33) read in our case

$$\{\rho_0 + p_0 - 2\phi_{tt} + 2\phi_t^2 + \tfrac{2}{3}\phi_t\theta_0\} \geqslant \lambda\phi_t\,|\sigma_0| , \quad (34)$$

$$\{\rho_0 - p_0 + 2\phi_{tt} + 4\phi_t + \tfrac{10}{3}\phi_t\theta_0\} \geqslant \lambda\phi_t\,|\sigma_0| , \quad (35)$$

where $\lambda = -2/\sqrt{3}, -1$ for $\gamma \neq 0$, respectively, and $\sigma_0 \equiv |2\underset{0}{\sigma}_{ab}\underset{0}{\sigma}_{ab}|^{1/2}$.

The equivalent conditions to be satisfied at the perfect-fluid stage are simply

$$\rho_0 + p_0 \geqslant 0 , \quad (36)$$

$$\rho_0 - p_0 \geqslant 0 , \quad (37)$$

and one can see by continuity that they are satisfied if (34) and (35) are; so one can take (36) and (37) as necessary conditions in order to impose (34) and (35).

IV. ISOMETRIES

So far we have made no assumptions about the possible isometry groups. In order to study the allowed possibilities let us recall the expressions (8) and (18) for the metric tensor. It is easy to see that they imply for the space-time, at least locally, a manifold product structure; i.e., $M = \mathbb{R} \times \Sigma_3$, where M designates the total space-time manifold and Σ_3 the three-dimensional hypersurfaces orthogonal to the velocity field of the fluid. Any occurring isometry group must then preserve this structure, that is, its orbits must be contained in Σ_3 (we discard the case of a timelike Killing vector), and the group can then be regarded as acting on three-dimensional manifold. Let us briefly examine the different possibilities.

As is well known, the maximal isometry group acting on a three-dimensional manifold is of order 6; such a group always admits a subgroup of order 3 acting transitively on three-dimensional orbits (this is the case of Friedman–Robertson–Walker models[1]). Five-dimensional isometry groups cannot occur on three-dimensional manifolds (Fubini's theorem[23]). The case of a group of order 4 is interesting in that it always admits a three-dimensional subgroup, giving rise then to two different situations: either the three-dimensional subgroup G_3 acts transitively on the three-dimensional hypersurfaces, or it acts on two-dimensional orbits. This last case constitutes, in the cosmological context, the so-called Kantowski–Sachs model.[33,34] In the case of a three-dimensional group G_3, there are again two possibilties, namely, G_3 acting transisitively on the three-dimensional hypersurfaces, and G_3 acting on two-dimensional (spacelike in the present case) orbits. In this case (G_3 on orbits S_2), the two-dimensional orbits must be of constant curvature, and the spatial part of the metric takes then one of the familiar forms listed in many reference books (see for instance, Ref. 23).

There are still three other possibile cases: a two-dimensional group G_2, a one-parameter group of motions G_1, and the case where no symmetries occur.

The cases in which the orbits of the group of isometries coincide with the three-dimensional hypersurfaces Σ_3 are called "spatially homogeneous models",[1] and they can be classified according to the Bianchi type of the three-dimensional group (or subgroup) occurring. These models exhibit many interesting characteristics from the physical and mathematical points of view,[1] and we shall take them as the framework for the example in the next section.

Notice that in our particular case the existing isometries are preserved throughout the whole process from viscous fluid to perfect fluid.

V. EXAMPLE: SPATIALLY HOMOGENEOUS BIANCHI TYPE III COSMOLOGIES

The purpose of this section is to give an example of how all the former considerations can be applied in a particular case. We shall take one of the families of exact perfect-fluid solutions of Bianchi type III given recently by Ram.[18] By means of a suitable choice of the function ϕ in Eq. (3) we shall construct a viscous-fluid solution that degenerates into the chosen perfect-fluid solution. The models given in Ref. 18 all correspond to geodesic, expanding ($\theta_0 > 0$), perfect-fluid solutions that satisfy the energy conditions (36) and (37), their line element being of the form[18]

$$\underset{0}{ds^2} = -dt^2 + A^2\,dx^2 + B^2 e^{2x}\,dy^2 + C^2\,dz^2 . \quad (38)$$

For $A = at + b$; $a,b \in \mathbb{R}$; Einstein's field equations imply

$$B = A = at + B , \quad (39)$$

$$C = c_1(at+b)^{\gamma} + c_2(at+b)^{-\gamma}, \quad |a| > 1 , \quad (40)$$

$$C = c_3\cos[\beta lg(at+b)] + c_4\sin[\beta lg(at+b)] , \quad |a| < 1 , \quad (41)$$

$$C = c_5 + c_6 lg(at+b), \quad |a| = 1 , \quad (42)$$

where $c_i \in \mathbb{R}$, $i = 1,...,6$, $\gamma \equiv (1/a)(a^2-1)^{1/2}$, and $\beta \equiv (1/a)(1-a^2)^{1/2}$.

In order to choose the function ϕ appearing in (3), we shall assume for the viscous fluid a relationship between viscosity and expansion of the type

$$\eta = -\lambda^2\theta + \gamma(t), \quad \lambda \in \mathbb{R}, \tag{43}$$

which for a suitable choice of the funciton $\gamma(t)$ implies

$$\phi_t = \beta^2 - \alpha^2\theta_0, \quad \alpha,\beta \in \mathbb{R}. \tag{44}$$

Equation (43), although completely general unless $\gamma(t)$ is specified, suggests that viscosity (i.e., some kind of friction between neighboring parts of the fluid) decreases as the fluid expands (i.e., as the parts of the fluid become more separated), which appears to be what one would expect.

From (44) one has

$$\phi_{tt} = -\alpha^2\dot{\theta}_0. \tag{45}$$

Therefore the matching hypersurface between the two metrics $\underset{0}{g}_{ab}$ and g_{ab} will be given by

$$t_1 \text{ such that } \dot{\theta}_0(t_1) = 0, \tag{46}$$

and from (44) and the requirements $\phi_t(t_1) = \phi(t_1) = 0$ one has

$$\beta^2/\alpha^2 = \theta_0(t_1), \tag{47}$$

$$\Omega(t) = \exp\left\{\beta^2 t - \alpha^2 \int^t \theta_0\, dt' + \phi_0\right\}, \tag{48}$$

with $\phi_0 \in \mathbb{R}$:

$$\phi_0 = \beta^2 t_1 - \int^{t_1} \theta_0\, dt. \tag{49}$$

Conditions (34) and (35) now read

$$(1-\alpha^2)\rho_0 + (1-3\alpha^2)p_0 + 2\beta^4 + 2\alpha^2(\alpha^2 - \tfrac{1}{3})\theta_0^2 + 2\beta^2(\tfrac{1}{3} - 2\alpha^2)\theta_0 - \alpha^2\underset{0}{\sigma}^2 \geqslant -\lambda(\beta^2 + \alpha^2\theta_0)\underset{0}{\sigma}, \tag{50}$$

$$(1+\alpha^2)\rho_0 + (3\alpha^2-1)p_0 + 4\beta^4 + 4\alpha^2(\alpha^2 - \tfrac{2}{3})\theta_0^2 + 2\beta^2(\tfrac{5}{3} - 4\alpha^2)\theta_0 + \alpha^2\sigma_0^2 \geqslant -\lambda(\beta^2 + \alpha^2\theta_0)\underset{0}{\sigma}, \tag{51}$$

where $\lambda = -2/\sqrt{3}, -1$ for $\gamma \neq 0$ and $\gamma = 0$, respectively.

Now, taking into account the expressions for p_0, ρ_0, and $\underset{0}{\sigma}$ for the perfect fluid described by the metric (38),[18] it is easy to see—after some straightforward calculations—that for a suitable choice of the parameters appearing in those expressions, (50) and (51) are satisfied for $-b/a < t \leqslant t_1$ [at $t = -b/a$, all the models described by (38)–(42) are singular].

One can now evaluate the entropy production density $\hat{s}$ given by Eq. (29). After a few calculations one has

$$\hat{s} = \tfrac{1}{3}(\alpha^2\theta_0 - \beta^2)\left(\frac{3a}{at+b} - \theta_0\right)^2 \times \exp\left\{-3\left[\beta^2 t - \alpha^2\int^t \theta_0\, dt' + \phi_0\right]\right\}, \tag{52}$$

for $|a| \neq 1$, and

$$\hat{s} = \frac{1}{12}(\alpha^2\theta_0 - \beta^2)\left(\frac{4a}{at+b} - \theta_0\right)^2 \times \exp\left\{-3\left[\beta^2 t - \alpha^2\int^t \theta_0\, dt' + \phi_0\right]\right\}, \tag{53}$$

for $|a| = 1$, the expansion θ_0 being in each case

$$\theta_0 = \frac{2a}{at+b} + \frac{a\gamma}{at+b}\left\{\frac{c_1(at+b)^\gamma - c_2(at+b)^{-\gamma}}{c_1(at+b)^\gamma + c_2(at+b)^{-\gamma}}\right\}, \quad |a| < 1, \tag{54a}$$

$$\theta_0 = \frac{2a}{at+b} + \frac{a\beta}{at+b} \times \left\{\frac{-c_3\sin[\log(at+b)] + c_4\cos[\log(at+b)]}{c_3\cos[\log(at+b)] + c_4\sin[\log(at+b)]}\right\}, \quad |a| < 1, \tag{54b}$$

$$\theta_0 = \frac{2a}{at+b} + \frac{2ac_6}{at+b}\{c_5 + c_6\log(at+b)\}^{-1}, \quad |a| = 1. \tag{54c}$$

Since the conformal factor (48) and the entropy production density (52)–(53) are continuous functions of the time coordinate, the implicit function theorem allows us to express one as a function of the other:

$$\Omega = \Omega(\hat{s}) = \begin{cases} \Omega(\hat{s}) \neq \text{const}, & \hat{s} > 0, \\ \Omega(\hat{s}) = 1, & \hat{s} = 0. \end{cases} \tag{55}$$

Therefore we can write

$$g_{ab} = \Omega^2(\hat{s})\underset{0}{g}_{ab}. \tag{56}$$

In other words, the metrics describing both stages of the fluid, nonperfect and perfect, are conformally related, and the conformal factor can be expressed as a function of the entropy production density $\hat{s}$ such that for positive entropy production ($\hat{s} > 0$) the fluid describing the material content of the space-time is a viscous fluid, and as the entropy production decreases to zero so does the viscosity (and, consequently, the anisotropy in the pressures), and the fluid becomes then a perfect one.

ACKNOWLEDGMENTS

I wish to thank Dr. G. S. Hall from the University of Aberdeen and Dr. L. Herrera from Universidad Central de Venezuela for many valuable discussions and suggestions.

The present work was supported by a British Council/M. E. C. Fleming Scholarship.

[1]M. P. Ryan and L. C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U.P., Princeton, NJ, 1975).

[2]G. Birkhoff, *Hydrodynamics, A Study in Logic, Fact and Similitude* (Princeton U.P., Princeton, NJ, 1960).

[3]G. L. Murphy, Phys. Rev. D **8**, 4231 (1973).

[4]V. A. Belinskii and I. M. Khalatnikov, Sov. Phys. JETP **42**, 205 (1975).

[5]V. A. Belinskii and I. M. Khalatnikov, Sov. Phys. JETP **45**, 1 (1977).

[6]B. L. Hu, *Cosmology of the Early Universe* (World Scientific, Singapore, 1984), p. 32.

[7]M. Novello and M. J. Rebouças, Astrophys. J. **225**, 719 (1978).

[8]D. Ray, J. Math. Phys. **21**, 2797 (1980).

[9]M. J. Rebouças and J. A. S. deLima, J. Math. Phys. **22**, 2699 (1981).

[10]M. J. Rebouças and J. A. S. de Lima, J. Math. Phys. **23**, 1547 (1982).

[11]J. M. Bradley and E. Svienstins, Gen. Relativ. Gravit. **16**, 1119 (1984).

[12]G. Mukharjee, J. Astrophys. Astron. **7**, 259 (1986).

[13]B. K. Nayak and B. K. Sahoo, Gen. Relativ. and Gravit. **21**, 211 (1989).
[14]S. Weinberg, Astrophys. J. **168**, 175 (1971).
[15]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Universe* (Cambridge U. P., Cambridge, 1973).
[16]R. Wald, *General Relativity* (University of Chicago, Chicago, 1984).
[17]Ch. A. Kolassis, N. O. Santos, and D. Tsoubelis, Class. Quantum Grav. **5**, 1329 (1988).
[18]S. Ram, J. Math. Phys. **30**, 757 (1989).
[19]G. S. Hall, Arab J. Sci. Eng. **9**, 87 (1984).
[20]G. F. R. Ellis, "Relativistic Cosmology," in *General Relativity and Gravitation*, Proceedings of the XLII Enrico Fermi Summer School, edited by R. K. Sachs (Academic, New York, 1971).
[21]G. S. Hall and A. D. Rendall, J. Math. Phys. **28**, 1837 (1987).
[22]G. S. Hall, "Curvature and Metric in General Relativity," in *Classical General Relativity* (Cambridge U.P., London, 1984).
[23]D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980).
[24]J. Carot and L. Mas, J. Math. Phys. **27**, 2336 (1986).
[25]C. Eckart, Phys. Rev. **58**, 919 (1940).
[26]L. Landau and E. M. Lifshitz, *Fluid Mechanics* (Addison–Wesley, Reading, MA, 1958).
[27]S. Weinberg, Astrophys. J. **168**, 175 (1971).
[28]W. Israel, Ann. Phys. (NY) **100**, 310 (1976).
[29]W. Israel and J. M. Stewart, Proc. R. Soc. London Ser. A **365**, 43 (1979).
[30]L. Lindblom and W. A. Hiscock, Astroph. J. **267** 384 (1983).
[31]W. A. Hiscock and L. Lindblom, Phys. Rev. D **31**, 725 (1985).
[32]S. K. Ma, *Statistical Mechanics* (World Scientific, Singapore, 1985).
[33]R. Kantowski, "Some relativistic cosmological models," Ph.D. thesis, University of Texas, Austin, 1966.
[34]R. Kantowski and R. K. Sachs, J. Math. Phys. **7**, 443 (1966).

Inflation in a spatially closed anisotropic universe

J. Ponce de León
Physics Department, University of Puerto Rico, Box AT, University Station, Río Piedras, Puerto Rico 00931

(Received 25 March 1988; accepted for publication 23 August 1989)

The effects of shear on the occurrence of inflation are studied on the basis of a simple model for a spatially closed universe which enters an inflationary era. It is assumed that the universe enters a vacuum-dominated phase in an abrupt transition that occurs everywhere at the same time. The space-time geometries, before and after the phase transition, are matched to each other via the Lichnerowicz junction conditions. The Einstein field equations are solved exactly for a viscous universe of the Kantowski–Sachs type. It is found that the inclusion of (positive) shear retards the occurrence of the vacuum phase transition. The magnitude of this effect depends on the mass of the universe at the time of the phase transition. For a universe with a mass of about 10 kg (which is a value usually associated with the mass of the region from which our universe originated), it is found that the inclusion of shear does not really have a large effect on the time at which the vacuum phase transition occurs. The generality of the results is also discussed.

I. INTRODUCTION

Since Guth's[1] discovery of the inflationary scenario for the early universe, a considerable amount of work has been devoted to the study of the problem of which subset of the initial data for the Einstein equations can undergo sufficient inflation to explain the present state of large scale homogeneity and isotropy of our universe.[2–12]

It has recently been shown[13] that initially expanding universes always undergo inflation provided that (i) the (positive) cosmological constant contributes to the energy–momentum tensor of the universe for all times, (ii) the universe has nonpositive spatial curvature, and (iii) the matter content of the universe satisfies the standard energy conditions.

One point of principle should be noted here, namely, that the energy density of the false vacuum only behaves like a cosmological constant during the finite period of time that the Higgs field ϕ spends in the "flat" region of the potential $V(\phi)$, evolving towards the true vacuum. In addition, the above result requires the discussion of two questions: first, the question of whether the matter in the dense stages of the very early universe satisfies the energy conditions; and, second, the question of the general conditions under which closed (positive spatial curvature) universes can undergo inflation.

The first question (or some variation of it) has been discussed in different contexts, viz., the avoidance of singularities,[14–19] the phenomenon of gravitational repulsion,[20] and in inflationary universe models.[9] Barrow[21] has recently argued that the assumption that the matter fields not driving inflation obey the strong energy condition is unsatisfactory, because the violation of this condition by one of the matter fields is a necessary condition for the occurrence of inflation (the cosmological constant arises from a massive scalar field that violates the strong energy condition).

Regarding the second question, the basic features of closed inflationary universes can be seen from the "initial-value constraint" equation

$$\Theta^2/3 = \Lambda + \tfrac{1}{2}\sigma_{ab}\sigma^{ab} + 8\pi T^0_0 - P/2\,, \qquad (1)$$

where Θ denotes the volume expansion and P is the spatial curvature [for details see Refs. 4 and 13]. It shows that if the positive curvature is dominant in the early universe, then it causes the universe to reverse its initial expansion, forcing it to recollapse before it can undergo inflation. If the effects of the positive curvature become dominant only asymptotically ($t \to \infty$), then it is clear that initially expanding closed universes will not recollapse, but they will continually expand. However, the fact that a closed universe is ever-expanding does not guarantee that it will enter an inflationary era. Indeed it has recently been shown, by the present author,[11–12] that there exist closed universes that are ever-expanding, but which have a non-de Sitter asymptotic behavior for large times.

The above discussion indicates that the following questions are of special interest when studying (closed) cosmological models.

(a) Does the matter content of the very early universe satisfy the energy conditions?

(b) How does the relaxation of the energy conditions affect the behavior of a cosmological model? Can it change the character of the singularities or of the general evolution of a cosmological model?

(c) Can the universe enter an inflationary epoch of exponential expansion? If it does, what is the time at which the phase transition occurs?

(d) What are the effects of shear on inflation in the case of anisotropic cosmologies? In particular, does the inclusion of shear retard or advance the occurrence of inflation?

In view of the complexity of the Einstein field equations the only way to investigate these questions is to examine the behavior of explicit cosmological models. The purpose of this paper is to present a simple model that allows one to investigate the above questions in the context of a closed inflationary universe.

This paper is organized as follows. In the next section, the model is described. In Sec. III, a specific class of solu-

 0022-2488/90/020371-07$03.00

tions of the field equations is presented, which is interpreted as a model for the universe before the phase transition. The physical properties of these solutions as well as the energy conditions are discussed in Sec. IV. In Sec. V, the question of whether the solutions of Sec. III can enter an inflationary era is discussed. In Sec. VI, some of the implications of our results are discussed.

II. DESCRIPTION OF THE MODEL

The model is based on the fact that, according to the new inflationary scenario,[22,23] the period of inflation occurs during the early stages of the "rollover" of the Higgs fields, while the energy density of the universe remains roughly constant. Therefore, I will assume that one can define a finite time t_p (the time of the "phase transition") such that the energy density of the universe does not significantly vary (is roughly constant) after t_p: any change in the energy density of the early universe is assumed to occur only before the time of the phase transition t_p.

It is to be noted that when one solves the field equations in the presence of a positive cosmological constant the phase transition to a vacuum-dominated phase (if it occurs) becomes completed only asymptotically in time. In our model the universe enters a vacuum-dominated phase in an abrupt transition that occurs everywhere at the same time t_p. This aspect of the model presents certain similarity with the supercooled phase transitions proposed few years ago by Hawking and Moss[24] for the exit from the inflationary stage without introducing inhomogeneities.

Since the matter content of the universe is assumed to be different before and after the phase transition at t_p, the space-time geometry (before and after t_p) is described by different solutions of the field equations. In order for our model to work, these solutions have to be matched across the separating spacelike hypersurface t_p via the Lichnerowicz junction conditions. This is an analogous treatment to that recently used by Wesson[25] to obtain a nonsingular cosmological model in which matter is produced from empty Minkowski space.

The change of the space-time metric will lead to the phenomenon of particle creation. I will disregard this effect by assuming that the energy density n of created particles is negligible compared to the energy density ρ_v of the false vacuum. This assumption is suggested by a recent work of Ford,[26] who showed (in another context) that n is typically of the order of $\rho_v^2/\rho_{\rm Pl}$, where $\rho_{\rm Pl}$ is the Plank energy density. For $\rho_v = (10^{14}\,{\rm GeV})^4$ and $\rho_{\rm Pl} = (10^{19}\,{\rm GeV})^4$, $n \sim 10^{-20}\rho_v$. Consequently, after the transition at t_p the energy density of the universe will be taken equal to ρ_v.

For a given space-time geometry, our method leads to a set of algebraic equations that relate the shear, the mass, and the size of the universe to the time t_p of the phase transition. Thus the solution of these equations allows one to investigate the questions (a)–(d) noted in the Introduction.

I have studied different spatially closed cosmologies starting from a line element of the Tolman–Bondi type. I found that generally (in view of the complexity of the solutions involved) the junction conditions lead to equations that cannot be analytically solved.

Therefore, in this work I confine my discussion to a specific class of closed cosmologies, namely, the Kantowski–Sachs (KS) cosmologies, for which all the equations can be solved exactly. KS cosmologies can be considered as particular (homogeneous) cases of the spatially closed Tolman–Bondi metrics in the case where the azimuthal metric coefficient depends on time only. Therefore, one can expect KS universes to contain the essential physics of more complicated (realistic) closed universes with shear and without vorticity. In addition, it may be worth mentioning that it has recently been argued that KS universes can be relevant to the description of phase changes in the early universe.[27–29]

III. A MODEL FOR THE UNIVERSE BEFORE THE PHASE TRANSITION

A. Equation of state and geometry

It is assumed that the early universe had positive spatial curvature and shear, and that before the time t_p of the vacuum phase transition it was filled with a uniform fluid with energy density ρ and pressure p, related to ρ by the equation of state

$$p = n\rho\,, \tag{2}$$

with

$$0 \leqslant n \leqslant 1\,. \tag{3}$$

In addition, the effects of viscosity are also taken into account. Therefore, the stress–energy tensor is taken as

$$T_{\mu\nu} = (\rho + p - \zeta\Theta)U_\mu U_\nu - (p - \zeta\Theta)g_{\mu\nu} + 2\eta\sigma_{\mu\nu}\,, \tag{4}$$

where U^μ is the fluid four-velocity, Θ is the expansion, $\sigma_{\mu\nu}$ are the components of the shear tensor, and ζ and η are the coefficients of bulk and shear viscosity, respectively.

There are several motivations for the introduction of viscosity in our model.

First, bulk viscosity may arise in different contexts during the evolution of the early universe, e.g., in the evolution of cosmic strings[30] due to their interaction with each other and with the surrounding matter, in a classical description of the (quantum) particle-production phases,[31] and in many other physical phenomena.[32]

Second, since it is assumed that the universe is shear anisotropic, dissipative processes due to shear viscosity are consistent with the model. Therefore, if some kind of dissipation occurred in the early universe, then it is reasonable to expect that it could have been associated not only with the bulk viscosity, but also with the shear viscosity. Therefore, in general, it is assumed here that ζ and η are different from zero.

Third, it will be shown in Sec. IV that, unless both bulk and shear viscosity are introduced in the model, the physical requirement of positiveness of the thermodynamic pressure is, in general, incompatible with the specific solutions we will discuss below.

Next, it is assumed that the space-time geometry can be described by a line element of the Kantowski–Sachs type, viz.,

$$ds^2 = dt^2 - e^{\lambda(t)}\,dr^2 - R^2(t)[d\theta^2 + \sin^2\theta\, d\phi^2]\,. \tag{5}$$

The Einstein field equations corresponding to this line element are as follows ($c=G=1$)[33]:

$$8\pi T^0_0=\frac{\dot\lambda\dot R}{R}+\frac{\dot R^2}{R^2}+\frac{1}{R^2}, \tag{6}$$

$$8\pi T^1_1=\frac{2\ddot R}{R}+\frac{\dot R^2}{R^2}+\frac{1}{R^2}, \tag{7}$$

$$8\pi T^2_2=8\pi T^3_3=\frac{\dot\lambda\dot R}{2R}+\frac{\ddot\lambda}{2}+\frac{\dot\lambda^2}{4}+\frac{\ddot R}{R}, \tag{8}$$

$$T_{\mu\nu}=0, \quad \text{for } \mu\neq\nu, \tag{9}$$

where $(0,1,2,3)\equiv(t,r,\theta,\phi)$ and the overdots denote a derivative with respect to t.

In the comoving coordinate system

$$U^\mu=(1,0,0,0), \tag{10}$$

the expansion and the nonvanishing components of the shear are given by

$$\Theta=\dot\lambda/2+2\dot R/R, \tag{11}$$

$$\sigma^1_1=-2\sigma^2_2=-2\sigma^3_3=-\tfrac{1}{3}(2\dot R/R-\dot\lambda). \tag{12}$$

Substituting Eqs. (2) and (10)–(12) into (4), one obtains

$$\rho=T^0_0, \quad p=nT^0_0, \tag{13}$$

$$\eta=(T^2_2-T^1_1)(2\dot R/R-\dot\lambda)^{-1}, \tag{14}$$

$$\zeta=[nT^0_0+\tfrac{1}{3}(T^1_1+2T^2_2)](\dot\lambda/2+2\dot R/R)^{-1}. \tag{15}$$

Thus there are five unknowns and only three equations. Therefore, in order to obtain specific solutions one has to make some additional assumptions.

B. A class of ever-expanding universes of positive curvature: Self-similar solutions to (13)–(15)

As we noted in the Introduction, Eq. (1) implies that there is a set of closed (positive spatial curvature) universes that recollapse. However, for the discussion of the questions (c) and (d) one needs to obtain expanding closed models that can cool down to the temperature of the inflationary phase transition. The question is then: how can one complete the field equations in order to obtain the desired kind of solutions?

Now I proceed to derive the solutions to the above equations which follow from the assumption that the quantities occurring in $T_{\mu\nu}$ are self-similar. This assumption is motivated by the fact that it excludes at the outset the possibility of having solutions representing recollapsing universes. In fact, following the classical notion of similarity, self-similar solutions (of the first kind) may arise only in systems without temporal or spatial characteristic scales.[34–42] On the other hand, recollapsing universes have intrinsic scale restrictions, namely, the age of the universe (i.e., the time needed for a full cycle), and the "radius" of the universe. Therefore, our assumption assures that our model universe solution is ever-expanding (or ever-contracting).[29]

From a mathematical viewpoint the assumption of self-similarity means that the space-time admits a homothetic Killing vector, viz.,

$$\underset{\xi}{L}\, g_{\mu\nu}=2g_{\mu\nu}, \tag{16}$$

where the left-hand side is the Lie derivative of the metric tensor with respect to the vector field ξ^μ.

In the case under consideration, by virtue of the spherical symmetry, one can set $\xi^2=\xi^3=0$, without loss of generality. Therefore, Eq. (16) reduces to

$$\xi^0{}_{,0}=1, \tag{17}$$

$$\dot\lambda\xi^0+2\xi^1{}_{,1}=2, \tag{18}$$

$$\dot R\xi^0=R \tag{19}$$

$$\xi^0{}_{,1}-e^\lambda\xi^1{}_{,0}=0, \tag{20}$$

where the commas denote partial derivatives. Since R is a function of t only, it follows from Eqs. (17) and (19) that

$$\xi^0=(t-t_0), \tag{21}$$

$$R=(t-t_0)/\alpha, \tag{22}$$

where t_0 and α are constants of integration. Note that by changing the origin of time one can set $t_0=0$, without loss of generality. Therefore, hereafter $t_0=0$.

Since e^λ and ξ^0 are functions of t only, it follows from Eqs. (20) and (18) that

$$\xi^1=\beta r, \tag{23}$$

where β is a separation constant. Consequently, from (18),

$$e^\lambda=C^2t^{2(1-\beta)}, \tag{24}$$

where C^2 is a constant of integration.

Thus, in summary, we have found that the sole assumption of self-similarity defines entirely the metric functions in Eq. (5). Now the substitution of Eqs. (22) and (24) into (6)–(9) and (13)–(15) gives the final form of the solution, as follows:

$$ds^2=dt^2-C^2t^{2(1-\beta)}dr^2-(t^2/\alpha^2)[d\theta^2+\sin^2\theta\, d\phi^2], \tag{25}$$

$$\rho=(3-2\beta+\alpha^2)/8\pi t^2, \tag{26}$$

$$p=n\rho, \quad 0\leqslant n\leqslant 1, \tag{27}$$

$$\zeta=\left[\frac{(\alpha^2+3)(1+3n)+2\beta^2-2\beta(2+3n)}{24\pi(3-\beta)}\right]\frac{1}{t}, \tag{28}$$

$$\eta=\frac{(\beta^2-2\beta-\alpha^2)}{16\pi\beta}\,\frac{1}{t}. \tag{29}$$

The nonvanishing components of the generator of the homothetic symmetry are given by Eqs. (21) and (23). The expansion and shear are

$$\Theta=(3-\beta)/t, \quad \sigma=-\beta/\sqrt{3}\,t. \tag{30}$$

The above solution represents a self-similar viscous universe of the Kantowski–Sachs type that is ever-expanding (ever-contracting) for $\beta<3$ ($\beta>3$) and $t>0$. For $\beta=3$, $\Theta=0$ and consequently there are no effects of dissipation due to bulk viscosity. In this case the universe is "static" in the sense that its volume is constant in time. However, the matter distribution, the "thickness" of the universe, as well as its extension in the two other directions change with time.

It should be noted at this point that the metric (25) is the only line element that has the property of self-similarity, in the case of universes of the Kantowski–Sachs type.

Furthermore, to the best of my knowledge, the simple viscous-fluid model given by Eqs. (25)–(29) has not appeared before in the literature.

IV. PROPERTIES OF THE MATTER DISTRIBUTION

In this section we will discuss the questions (a) and (b) noted in the Introduction. I will show that the viscous-fluid model given by Eqs. (25)–(29) is consistent with the physical requirements.

A. Energy conditions and singularities

The metric (25) contains two free parameters, viz., α and β, which can be used in such a way as to assure that it constitutes a physically acceptable solution to the field equations. Recall that, according to Hawking and Ellis,[43] a solution of the Einstein field equations is physically acceptable if the components of the energy–momentum tensor satisfy at least one of the standard energy conditions. For the distribution under consideration these conditions read as follows[44]: The "weak" energy condition requires $\rho\geqslant 0$ and $(\rho + p_i)\geqslant 0$, where the p_i represent the principal pressures; the "dominant" energy condition requires $\rho\geqslant 0$ and $-\rho\leqslant p_i\leqslant\rho$; and, finally, the "strong" energy condition requires $(\rho + p_i)\geqslant 0$ and $(\rho + \Sigma p_i)\geqslant 0$.

It is easy to verify that the weak and the dominant energy conditions are satisfied for all values of α provided that

$$-\sqrt{2+\alpha^2}\leqslant\beta\leqslant 1 . \tag{31}$$

The fulfillment of the strong energy condition provides a more stringent lower limit on β than occurs in (31), viz.,

$$0\leqslant\beta\leqslant 1 . \tag{32}$$

Note the relation between the above conditions and the singularities: If all the energy conditions are satisfied, then the singularity is either pointlike ($\beta<1$) or barrel-like ($\beta = 1$); if neither the dominant nor the strong energy conditions are satisfied (but the energy density is positive), then the singularity is cigarlike.

B. Conditions on the physical quantities

In general the metric (25) represents an anisotropic distribution of matter, in the sense that for arbitrary values of β the principal pressures p_i are unequal. In fact,

$$p_1 = -T^1_1 = p - \zeta\Theta - 2\eta\sigma^1_1 = -(1+\alpha^2)/8\pi t^2 , \tag{33}$$

$$p_2 = p_3 = -T^2_2 = -T^3_3 = p - \zeta\Theta - 2\eta\sigma^2_2 = -(1-\beta)^2/8\pi t^2 . \tag{34}$$

We see that (i) the p_i are negative for all values of α and β, and (ii) $p_1 = p_2 = p_3$, for $\beta = 1 \pm (1+\alpha)^{1/2}$ only. For other values of β the source is anisotropic. Note also that, according to (31) and (32), only the choice of the negative sign leads to "isotropic" distributions with acceptable physical properties.

Equation (29) shows that η diverges for $\beta = 0$. This is not surprising, because for this value the fluid is shear-free and consequently the introduction of shear viscosity is inappropriate. Therefore, in what follows, it will be assumed that $\beta \neq 0$.

In order to ensure that the viscous fluid interpretation of the solution is acceptable, the physical quantities in (26)–(29) must behave in a satisfactory manner, viz., $\rho\geqslant 0$, $p\geqslant 0$, $\rho\geqslant p$, $\zeta\geqslant 0$, and $\eta\geqslant 0$. It is easy to verify that these conditions are satisfied in the range

$$(1-\sqrt{1+\alpha^2})\leqslant\beta<0 . \tag{35}$$

In this range the strong energy condition is violated, but the weak and dominant energy conditions are satisfied. Thus, in summary, we have found that the KS viscous-fluid universe given by Eqs. (25)–(29) is physically acceptable, but it necessarily violates the strong energy condition.

C. Effects of viscosity

Equations (33) and (34) clearly show the effects produced by the shear and bulk viscosity.

In fact, it follows from these equations that the matter content of the universe, in our model, can satisfactorily be interpreted as a "normal" fluid (i.e., as a fluid whose thermodynamic pressure p is equal in all directions) only due to the introduction of shear viscosity. If there were no shear stresses during the expansion of the fluid the pressure p would not obey Pascal's principle.

The introduction of bulk viscosity assures that the thermodynamic pressure p is positive. In other words, bulk viscosity allows the existence of solutions that violate the strong energy condition, but that satisfy a reasonable equation of state, viz., $p = n\rho$ with $0\leqslant n\leqslant 1$. In the absence of bulk viscosity the existence of this kind of solutions is, of course, not possible.

D. Self-similarity, physical meaning of α and β

As we discussed at the beginning of Sec. III B the assumption of self-similarity excludes recollapsing solutions. Therefore, although the Kantowski–Sachs universes have positive spatial curvature, the model under consideration represents universes that expand ($t>0$) or contract ($t<0$) forever, without dimensional constraints. Indeed, setting

$$r = [\alpha^{(\beta-1)}/C\beta]\,\bar{r}^{\beta} \tag{36a}$$

renders the metric (25) in a manifest scale-free form, viz.,

$$ds^2 = dt^2 - \xi^{2(1-\beta)}\,d\bar{r}^2 - \bar{r}^2\xi^2\,d\Omega^2 , \tag{36b}$$

where the similarity variable ξ was defined as

$$\xi = t/\alpha\bar{r} . \tag{37}$$

In the new coordinate $\bar{r}$ the generator of the homothetic symmetry is given by

$$\xi^{\mu} = (t,\bar{r},0,0) . \tag{38}$$

From the fact that the physical requirements discussed in Secs. IV A and IV B do not impose any restriction upon α, one could ask whether one can simply set $\alpha = 1$. However, Eq. (36) shows that α cannot be eliminated by means of coordinate transformations, which indicates that this parameter represents some physically meaningful quantity.

In order to discover the physical meaning of α, let us introduce the mass function $M(r,t)$ of Misner and Sharp,[45] viz.,

$$M(r,t) = (R/2)\{1 + e^{-\nu}\dot{R}^2 - e^{-\lambda}R'^2\}, \quad (39)$$

where e^ν stands g_{00}, a prime denotes a partial derivative with respect to r, and $M(r,t)$ is interpreted as the total mass or energy inside a sphere of circumference $2\pi R$. For the line element (25), the above expression reduces to

$$M = [(\alpha^2 + 1)/2\alpha^3]\, t. \quad (40)$$

Consequently, the parameter α measures the total mass of the universe at a given time t. Similarly, Eq. (30) indicates that β measures the expansion and shear of the universe at a given time.

V. TRANSITION TO AN INFLATIONARY ERA

In this section we will investigate the questions (c) and (d) noted in the Introduction, namely, whether our ever-expanding viscous-fluid universe can enter an inflationary era. We will see that the parameters α and β determine the time of the phase transition t_p.

For $t > 0$ ($\alpha > 0$) the viscous-fluid universe (25)–(29) will continually expand and cool. It would then cool down to the temperature of the GUT phase transition and, according to the inflationary universe models, the universe would undergo extreme supercooling, approaching not the true vacuum, but the false-vacuum state with a constant positive energy density ρ_v. In such a state the energy-momentum tensor in Eqs. (6)–(9) takes the form

$$T_{\mu\nu} = \rho_v g_{\mu\nu}. \quad (41)$$

The corresponding solution to the field equations is given by[11]

$$e^\lambda = A^2\dot{R}^2, \quad (42)$$

$$\dot{R}^2 = \Lambda R^2/3 + D/R - 1, \quad (43)$$

where A and D are constants of integration, and Λ is the usual cosmological constant $\Lambda = 8\pi\rho_v$.

In accordance with the discussion of Sec. II, it is now assumed that one can define a finite time t_p after which the energy density of the universe will be given by Eq. (41). In order for this assumption to be consistent with the field equations one has to match the geometry of the space-time before t_p with the geometry of the space-time after t_p.

Recall that two regions of the space-time are said to match across a separating hypersurface (say S) if the metric tensor and all its first-order partial derivatives are continuous across S (Lichnerowicz junction conditions).

A. The matching

In the case under consideration the separating hypersurface S is defined by

$$t - t_p = 0. \quad (44)$$

Because of the simplicity of the metrics (25), (42), and (43), the junction conditions can be solved exactly. In fact, the continuity of $\dot{R}$ at t_p gives

$$\Lambda t_p^3 - 3(\alpha^2 + 1)t_p + 3D\alpha^3 = 0. \quad (45)$$

The continuity of $\dot{\lambda}$ at t_p gives

$$2\Lambda t_p^3 - 6(1-\beta)t_p - 3D\alpha^3 = 0. \quad (46)$$

Solving these two equations we obtain

$$t_p = (1/\sqrt{\Lambda})(3 + \alpha^2 - 2\beta)^{1/2}, \quad (47)$$

$$D = (2/3\sqrt{\Lambda})[(\alpha^2 + \beta)/\alpha^3](3 + \alpha^2 - 2\beta)^{1/2}. \quad (48)$$

From the continuity of the metric functions at t_p, one obtains

$$R_p \equiv R(t_p) = (1/\alpha\sqrt{\Lambda})(3 + \alpha^2 - 2\beta)^{1/2}, \quad (49)$$

$$(A/C)^2 = \alpha^2[(3 + \alpha^2 - 2\beta)/\Lambda]^{(1-\beta)}. \quad (50)$$

The above equations show that the assumed phase transition is consistent with the matching conditions. Indeed the viscous-fluid universe under consideration has acceptable physical properties in the range $\alpha > 0$, $(1 - \sqrt{1+\alpha^2}) \leqslant \beta < 0$ [Eq. (35)]. It is not difficult to see that, within this range, t_p, D, R_p, and $(A/C)^2$ are positive and real quantities. Therefore, for every α and β (belonging to the physically allowed range), one can find from Eqs. (47) and (48) the corresponding values of t_p and D for which the solution (25)–(29) enters a vacuum-dominated phase. Thus the only condition required for the correct matching of both solutions is that $D > 0$.

Consequently, the question of whether the viscous-fluid universe can undergo inflation reduces to the investigation of the conditions under which the false vacuum metric (42) and (43), with $D > 0$, has inflationary solutions.

B. Range of parameters that allow inflation

Obviously, the solution will not undergo inflation for all positive values of D. Some features of the solutions of Eq. (43) may be exhibited by introducing the auxiliary function $V(R)$ by

$$V(R) = 1/R^2 - D/R^3, \quad (51)$$

in terms of which Eq. (43) reads

$$\dot{R}^2 = R^2[\Lambda/3 - V(R)]. \quad (52)$$

Thus the region of allowed values of R is given by the inequality

$$V(R) \leqslant \Lambda/3. \quad (53)$$

Figure 1 shows that there are three different types of solutions, which are marked with the roman numerals I–III. The solutions marked with I are recollapsing solutions. Regions II and III represent the solutions that may undergo inflation, because for these solutions R may take arbitrary large values.

It is easy to prove that our viscous-fluid solutions are matched to the false vacuum solutions marked with III, in the figure, and not to those of I or II. In fact, the solutions of III are those for which $V_{max} < \Lambda/3$, where V_{max} denotes the maximum value of $V(R)$, which is $V_{max} = (4/27D^2)$. Thus $D > 2/3\sqrt{\Lambda}$ in III, while, for the solutions of I and II, $0 < D < 2/3\sqrt{\Lambda}$. Examination of Eq. (48) reveals that

$$\varepsilon \equiv [(\alpha^2 + \beta)/\alpha^3](3 + \alpha^2 - 2\beta)^{1/2} > 1, \quad (54)$$

for all values of α and β in the range of allowed values given by Eq. (35). Consequently, in our model, $D > 2/3\sqrt{\Lambda}$ and so

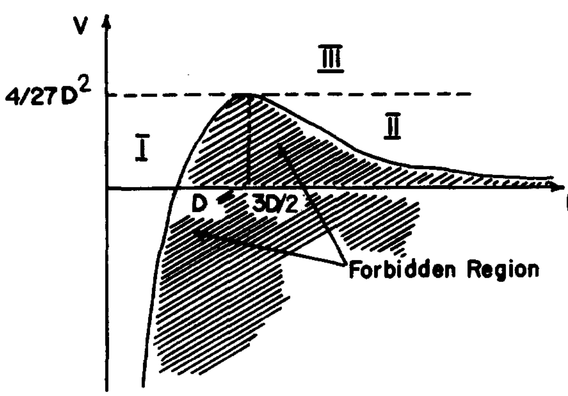


FIG. 1. Shape of the auxiliary function $V(R)$ defined by Eq. (51). The figure shows that Eq. (43) has three types of solutions. The solutions corresponding to I are recollapsing, while those of II and III may produce an inflationary scenario. The viscous fluid solution (25)–(29) is matched to the inflationary solutions of III ($\Lambda/3 > V_{max} = 4/27D^2$) for all allowed values of α and β.

after the transition at $t = t_p$ the universe is described by the vacuum solutions of III.

The conclusion is that our model undergoes inflation for all values of α and β allowed by Eq. (35).

C. Asymptotic behavior

We now proceed to discuss question (c) quoted in the Introduction.

Equation (43) can be analytically integrated only for one specific value of D, namely, $D = 2/3\sqrt{\Lambda}$ ($V_{max} = \Lambda/3$), which corresponds to a recollapsing solution[11] (there is another analytic solution[46] for $D = 0$, but according to the preceding discussion it cannot be matched to our viscous-fluid universe model). For all other positive values of D the solution is given in terms of elliptic integrals. However, the essential features of the solutions may be seen from the expressions for the expansion and shear, viz.,

$$\Theta = \sqrt{3\Lambda}\left(1 - \frac{2}{x^2} + \frac{\varepsilon}{x^3}\right)\left(1 - \frac{3}{x^2} + \frac{2\varepsilon}{x^3}\right)^{-1/2}, \quad (55)$$

$$\sigma = \frac{\sqrt{\Lambda}}{x^2}\left(1 - \frac{\varepsilon}{x}\right)\left(1 - \frac{3}{x^2} + \frac{2\varepsilon}{x^3}\right)^{-1/2}, \quad (56)$$

where ε is the parameter defined in Eq. (54) and $x \equiv \sqrt{\Lambda}R$. These equations (with $\varepsilon > 1$) clearly show that the universe, after the transition to a vacuum-dominated phase, will rapidly evolve to an isotropic state with an expansion rate $\Theta = \sqrt{3\Lambda}$. As the universe expands the term $(D/R - 1)$ in Eq. (43) becomes small compared with $\Lambda R^2/3$. This will lead to an exponential expansion of R and the universe will rapidly approach a state locally indistinguishable from the de Sitter one, viz.,

$$ds^2 \simeq dt^2 - e^{2\sqrt{\Lambda/3}\,t}\left[dr^2 + r_0^2\, d\Omega^2\right], \quad (57)$$

where we have set $A^2 = 3/\Lambda r_0^2$, $\simeq$ means asymptotically equal, and r_0 is a constant of integration.

According to the inflationary universe models the false vacuum state is not stable, but at the end of inflation it would decay into particles with a subsequent reheating of the universe. This would end in a hot radiation-dominated homogeneous and isotropic universe. Thus our closed viscous-fluid universe can give rise to an acceptable inflationary scenario.

D. Effects of shear

We now proceed to discuss question (d) quoted in the Introduction. The shear and mass of the universe at the time t_p of the phase transition depend on the choice of α and β.

As an illustrative example, let us consider the case $\alpha = 1$ and $\beta = (1 - \sqrt{2})$, in which case the principal stresses of the energy–momentum tensor are equal to each other. From Eqs. (47), (40), and (30) we obtain $t_p \sim 1.26 \times 10^{-35}$ sec, $M_p \sim 5.1$ kg, and $\sigma_p \sim 1.88 \times 10^{34}$ sec^{-1}, where the value $\rho_v \sim 1.6 \times 10^{97}$ erg/cm^3 ($\sqrt{3/\Lambda} \sim 10^{-35}$ sec) has been used.

Note that, although, by Eq. (40), M depends explicitly only on α, the value of M_p (the mass of the universe at the time of the phase transition) depends also on β through t_p. The same thing occurs with the shear [Eq. (30)] evaluated at t_p, which implicitly depends on α. For a fixed M_p the possible values of the parameters α and β are given by

$$\beta = \tfrac{3}{2} + \alpha^2/2 - 6m^2\alpha^6/[4.05(\alpha^2 + 1)]^2,$$

$$1 - \sqrt{1 + \alpha^2} \leqslant \beta < 0, \quad (58)$$

where the dimensionless parameter m measures the mass M_p in kg, viz., $M = m$ kg. The solutions of (58) have been used to draw the effects of the shear on the time of the phase transition for various values of M_p. Figure 2 records these effects. It shows that [in the range given by (58)] the dependence between t_p and σ_p is (to a high degree of accuracy) lineal and that t_p increases with the increase of σ_p.

VI. SUMMARY AND CONCLUSION

Based upon the fact that inflation occurs while the energy density of the universe remains (roughly) constant, we have constructed here a model for an inflationary closed universe in which the transition to a vacuum-dominated state occurs at some finite time t_p after which the energy density of the universe is constant. The phase transition is described as a change in the space-time geometry. The universe makes a transition from the matter-dominated (preinflationary) state described by Eqs. (25)–(29) to the vacuum-dominated

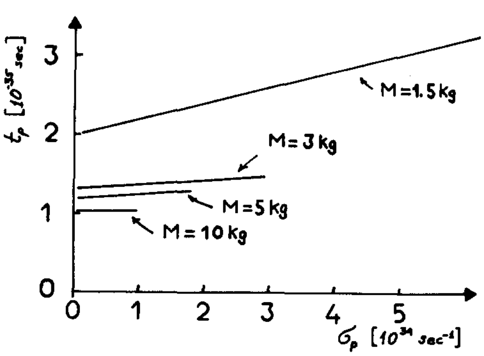


FIG. 2. Effects of shear and mass on the time of the phase transition t_p. The figure gives t_p [Eq. (47)] versus the shear [Eq. (30)] for various fixed values of M_p. The parameters α and β are the solutions of Eq. (58). It shows that t_p increases linearly (in the range under consideration) with the increase of σ_p.

(inflationary) state described by Eqs. (42) and (43). The strongest assumption is that this transition occurs everywhere at the same time, but it has been shown that this assumption is consistent with the usual continuity requirements (matching conditions). The main motivation for introducing this model has been the study of the effects caused by the shear and mass of the universe on the occurrence of the vacuum phase transition.

We have seen that the inclusion of (positive) shear retards the occurrence of the vacuum phase transition. This effect is, however, very weak because the inclusion of large shear of the order of 10^{34} sec^{-1} causes very little change in the time of occurrence of the phase transition. In addition, this effect depends on the mass M_p of the universe at the moment of the phase transition. In fact, Fig. 2 shows that for large masses, t_p is relatively insensitive to the change of σ. Only for "small" values of M_p does the "retardation effect" become manifest.

It is also interesting to note that there is a close connection between the mass of the universe and the upper limit of the shear at the time of the phase transition t_p. Massive universes inhibit the shear to take very large values and this causes these universes to enter a vacuum-dominated phase at the GUT time $t_G \sim \sqrt{3/\Lambda}$ (see Fig. 2).

Thus in the context of the models discussed here, one can conclude that the inclusion of shear does not really have a large effect on the phase transition. A very large (positive) shear can only lead to a significant retardation of the time of the occurrence of the phase transition, if the mass of the universe at that time were sufficiently small. Whether or not these results are still valid in other (spherically symmetric) anisotropic, spatially closed cosmologies is an interesting subject for further investigations. This is especially true, because our results are apparently in disagreement with other results in the literature, where it is argued that, if inflation occurs in some isotropic universe, then the addition of anisotropy can only make the inflation occur more easily and at earlier times.[9] In view of this disagreement, further investigation is needed for a better understanding of the role of the shear in the early universe.

[1]A. H. Guth, Phys. Rev. D **23**, 347 (1981).
[2]J. D. Barrow, in *The Very Early Universe*, edited by G. W. Gibbons, S. W. Hawking, and S. Siklos (Cambridge U.P., London, 1983), p. 267.
[3]W. Boucher and G. W. Gibbons, Ref. 2, p. 273.
[4]R. M. Wald, Phys. Rev. D **28**, 2118 (1983).
[5]G. Steigman and M. S. Turner, Phys. Lett. B **128**, 295 (1983).
[6]J. D. Barrow and J. Stein-Schabes, Phys. Lett. A **103**, 315 (1984).
[7]M. S. Turner and L. Widrow, Phys. Rev. Lett. **57**, 2237 (1986).
[8]L. Jensen and J. Stein-Schabes, Phys. Rev. D **34**, 931 (1986).
[9]J. D. Barrow, Phys. Lett. B **180**, 335 (1986); **183**, 285 (1987); **187**, 12 (1987).
[10]J. Stein-Schabes, Phys. Rev. D **35**, 2345 (1987).
[11]J. Ponce de León, Phys. Lett. A **126**, 75 (1988).
[12]J. Ponce de León, Phys. Lett. A **129**, 367 (1988).
[13]L. Jensen and J. Stein-Schabes, Phys. Rev. D **35**, 1146 (1987).
[14]F. J. Tipler, Phys. Rev. D **17**, 2521 (1978).
[15]T. A. Roman and P. G. Bergman, Phys. Rev. D **28**, 1265 (1982).
[16]J. D. Bekenstein, Phys. Rev. D **11**, 2072 (1975).
[17]G. L. Murphy, Phys. Rev. D **8**, 4231 (1973).
[18]P. N. Baaklini, Phys. Lett. A **66**, 357 (1978).
[19]L. Parker and S. A. Fulling, Phys. Rev. D **7**, 2357 (1973).
[20]J. Ponce de León, J. Math. Phys. **28**, 410 (1987); **29**, 197 (1988).
[21]J. D. Barrow, Nucl. Phys. B **296**, 697 (1988).
[22]A. D. Linde, Phys. Lett. B **108**, 389 (1982); **114**, 431 (1982).
[23]A. Albrecht and P. J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).
[24]S. W. Hawking and I. G. Moss, Phys. Lett. B **110**, 35 (1982).
[25]P. S. Wesson, Astron. Astrophys. **151**, 276 (1985).
[26]L. H. Ford, Phys. Rev. D. **35**, 2955 (1987).
[27]R. N. Henriksen, A. G. Emslie, and P. S. Wesson, Phys. Rev. D **27**, 1219 (1983).
[28]P. S. Wesson, Phys. Rev. D **34**, 3925 (1986).
[29]J. Ponce de León, J. Math. Phys. **29**, 2479 (1988).
[30]J. P. Ostriker, C. Thompson, and E. Witten, Phys. Lett. B **180**, 231 (1986).
[31]B. L. Hu, in *Advances in Astrophysics*, edited by L. Z. Fang and R. Ruffini (World Scientific, Singapore, 1983).
[32]S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), p. 593.
[33]R. Kantowski and R. K. Sachs, J. Math. Phys. **7**, 443 (1966).
[34]L. I. Sedov, *Similarity and Dimensional Methods in Mechanics* (Academic, New York, 1959).
[35]A. H. Cahill and M. E. Taub, Commun. Math. Phys. **21**, 1 (1971).
[36]G. I. Barenblatt and Ya. B. Zel'dovich, Ann. Rev. Fluid Mech. **4**, 285 (1972).
[37]B. J. Carr and S. W. Hawking, Mon. Not. R. Asron. Soc. **168**, 399 (1974).
[38]Ya. B. Zel'dovich, Ann. Rev. Fluid Mech. **9**, 215 (1977).
[39]R. N. Henriksen and P. S. Wesson, Astrophys. Space Sci. **53**, 429, 445 (1978).
[40]P. S. Wesson, Astrophys. J. **228**, 647 (1978).
[41]P. S. Wesson, Astron. Astrophys. **76**, 200 (1979); **80**, 296 (1979).
[42]A. A. Coley and B. O. J. Tupper, Astrophys. J. **288**, 418 (1985).
[43]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U.P., London, 1973), p. 117.
[44]See Ref. 43, p. 88.
[45]C. W. Misner and D. H. Sharp, Phys. Rev. B **136**, 571 (1964).
[46]Ø. Grøn, J. Math. Phys. **27**, 1490 (1986).

On the Legendre transformation for a class of nonregular higher-order Lagrangian field theories

Guido Magnano
S.I.S.S.A., Str. Costiera 11, 34014 Miramare Grignano, Trieste, Italy
Marco Ferraris
Dipartimento di Matematica, Università di Cagliari, Via Ospedale 72, 09124 Cagliari, Italy
Mauro Francaviglia
Istituto di Fisica Matematica "J.-L. Lagrange," Università di Torino, Via C. Alberto 10 10123 Torino, Italy

(Received 10 March 1987; accepted for publication 30 August 1989)

In the framework of higher-order calculus of variations, the generalized Legendre transformation for a wide class of Lagrangians is considered, which depend in a nonregular way on the derivatives of maximal order. A rigorous theory is discussed for Lagrangians depending on a constant rank set of affine combinations of these derivatives. This allows the reduction of the Poincaré–Cartan formalism and the Hamiltonian formalism to the appropriate constraint in the appropriate phase space of the problem. The case considered here covers many important physical examples, such as the Yang–Mills theories (at order one) and relativistic metric theories of gravitation (at order two).

I. INTRODUCTION

In recent years, it has been noticed by many authors that some higher-order nonlinear Lagrangians can be transformed by a suitable change of dynamical variables into simpler ones, possibly linear and of lower order. This phenomenon occurs in different contexts. To our knowledge, it has been used in the following situations.

(a) By Einstein and Eddington,[1,2] and later by Schrödinger,[3] in view of possible applications to unified theories, to establish a link between "metric" and "affine" formulations of general relativity and other gravitational models. This idea (which is described in Sec. V A) was given renewed attention some years ago, especially by Kijowski and one of us[4,5] to prove the dynamical equivalence between a whole class of "purely affine" gravitational theories and general relativity (possibly coupled with external fields), and to derive a unified formulation of Einstein–Maxwell equations from an affine Lagrangian.[6,7]

(b) By Higgs in 1959,[8] and more recently by Whitt,[9] in a second-order context, to prove the dynamical equivalence with general relativity (plus suitable matter) for a class of gravitational theories in metric formulation (namely, those based on quadratic Lagrangians of the type $L=\sqrt{-g}\,[aR+bR^2]$). Higgs also considered the case of Lagrangians quadratic in the Ricci tensor $R_{\mu\nu}$. Recently, we were able to show[10,11] that this equivalence holds for a wider class of Lagrangians, containing at least all covariant Lagrangians depending arbitrarily on $R_{\mu\nu}$. In particular, our results allow us to recover and extend earlier work by Stelle,[12] who showed that, in the case $L=\sqrt{-g}\,(aR+bR^2+cR_{\mu\nu}R^{\mu\nu})$, the particle spectrum contains a spin-two massless graviton satisfying Einstein's equations, coupled with a spin-two ghost and possibly a scalar field. Nontrivial generalizations to supergravity of these earlier results can be found in recent literature.[13]

All these examples refer to gravitational field theories, for which the investigation of such equivalence problems has been carried out more deeply. This is easily explained if one considers that, in general, higher-order Lagrangians are commonly neglected for well-known reasons (for instance, the lack of positivity of the metric structure in the space of states, after quantization), while in purely metric gravitational theories the appearance of second-order derivatives is unavoidable for covariance reasons. However, in dealing with the gravitational field, the transformation methods proposed by the aforementioned authors are commonly believed to rely only on the peculiar features of the dynamical variables (metric tensor and/or linear connection), so that the very nature and the general validity of these procedures are somehow hidden. For instance, it seems that the well-known equivalence between some quadratic metric theories and general relativity is apparently considered in the current literature as a purely accidental feature.

On the other hand, as it was earlier suggested by Schrödinger,[3] the method thereby proposed, rather than being an ad hoc prescription restricted to gravitational theories only, should be a particular application of a general transformation rule suitable to deal with any (sufficiently regular) Lagrangian field theory. The nature of the Legendre-type transformation of this method seemed to be intuitively evident; however, the construction of a consistent mathematical framework justifying this terminology also for higher-order nonhyperregular models, such as those for which the momenta are not all independent, was until now an open problem. For instance, we might mention a remarkable series of papers by Kuchař[14] where, among other things, an explicit prescription for the Legendre transformation of any field theory was envisaged. However, we stress that the arguments of Ref. 14, although interesting, constitute only a first preliminary step toward solving the problem; in fact, they refer to the first-order case only, depend on regularity as-

 0022-2488/90/020378-10$03.00

sumptions that are not well specified, and are of essentially local character.

We shall present here a rigorous and global description of the generalized Legendre transformation and the Hamiltonian formulation for higher-order field theories, under regularity conditions that are weak enough to cover all of the well-known examples of physical interest. To avoid any possible confusion about our terminology, let us briefly recall that the Hamiltonian picture of dynamics, which is essentially unique for mechanical systems, admits different generalizations to the field-theoretical domain. Starting from the Lagrangian formulation of field theory, which is governed by a variational principle, we can obtain an equivalent description of dynamics in terms of energy and energy flows. This is in fact the description which in most of the physical literature is assumed to be the Hamiltonian one, and which generates the useful dynamical splittings commonly known as ADM's formalisms (see, e.g., Refs. 14–19 and references quoted therein).

On the other hand, geometric approaches to the Legendre transformation,[20] and a better understanding of the role in which the Poincaré–Cartan form plays in both mechanics and field theories,[21–24] have led in recent years to a different generalization of the Hamiltonian formalism. Roughly speaking, this different perspective is based on the identification of the canonical momenta with the components of the full differential of the Lagrangian density, rather than the components conjugated to the time derivatives only (whereby the time variable is either predefined or somehow selected on physical grounds).

A general formulation of the global Legendre transformation theory for higher-order variational principles was already described in the hyperregular case, in terms of morphisms between suitable jet bundles.[25,26] In this formalism a predominant role was assigned to the so-called Legendre bundle. Unfortunately, physically meaningful theories cannot be hyperregular in the standard sense, mainly as a consequence of covariance or symmetry requirements. A considerable step forward, aimed at extending the domain of applications of this formalism, therefore needs to investigate weaker regularity conditions (see also Ref. 27 for the notion of regularity). To fix the ideas, we recall that whenever the Lagrangian density is not regular, the image of the Legendre map cannot be the whole Legendre bundle; as a matter of fact, it is precisely in this image that a correct Hamiltonian formulation may be given.

In this paper we shall consider Lagrangian theories of any order, assuming explicitly that the Lagrangian depends on the highest-order derivatives through a set of linear combinations of arbitrary, but not maximal, rank. This situation, although very particular from the mathematical viewpoint, is very common in physical models. To make the paper self-contained, Secs. II and III will be mainly devoted to a review of the notation and on the theory of the Legendre transformation for regular cases. We shall then produce a suitable Hamiltonian description on the image of the Legendre map, whereby the full set of (generally redundant) highest-order momenta is from the beginning replaced by suitably defined momentumlike new variables, whose definition is directly suggested by the explicit assumptions made on the Lagrangian. Various examples will be considered at the end of this paper. In this way we hope to help a better mathematical understanding of all the gravitation-theoretic examples mentioned above, by showing the relation between those particular cases and the general framework presented here.

A more direct Legendre transformation method for first-order theories will be discussed elsewhere by one of us,[28] under much weaker regularity conditions, providing a Hamiltonian description that is based on expressing all highest-order momenta in terms of an arbitrary set of coordinates in the image of the Legendre map.

II. PRELIMINARIES AND NOTATION

We assume that the reader is familiar with standard concepts and notation of differential geometry on fibered manifolds and jet prolongation theory.[23,29]

We shall deal with variational principles, defined in terms of the following basic objects.

(a) A differentiable manifold $\mathbf{X}$, $\dim(\mathbf{X}) = m$, representing the space of physical parameters (for instance, a four-dimensional space-time); a local coordinate system on $\mathbf{W}\subset\mathbf{X}$ will be denoted by $(\mathbf{W},x^\lambda)$.

(b) A fibered manifold $\mathbf{Y}$ over $\mathbf{X}$, which is interpreted as the configuration space of the model; its (local) sections $\sigma\in\Gamma(\mathbf{Y})$ will represent the physical fields, or dynamical variables. We shall use only fibered parametrizations of $\mathbf{Y}$, denoted by $(\mathbf{U},x^\lambda,y^i)$, $1\leqslant i\leqslant h$ [$h=\dim(\mathbf{Y})-m$ is the dimension of the fiber of $\mathbf{Y}$, and $\mathbf{U}$ is an open subset of $\mathbf{Y}$ projecting over the domain $\mathbf{W}$ of a chart $(\mathbf{W},x^\lambda)$ of $\mathbf{X}$]. The rth order jet prolongation of $\mathbf{Y}$ will be denoted by $\mathbf{J}^r\mathbf{Y}$. The "natural fibered chart" on $\mathbf{J}^r\mathbf{Y}$, induced by the chart $(\mathbf{U},x^\lambda,y^i)$ on $\mathbf{Y}$, will be denoted by $(\mathbf{J}^r\mathbf{U},x^\lambda,y^i_{\underline{\nu}})$, $|\underline{\nu}|\leqslant r$. Multi-indices will be denoted by underlined Greek letters: $\underline{\nu}\equiv(\nu_1,\ldots,\nu_m)\in\mathbf{N}^m$. We set

$$|\underline{\nu}|\equiv\nu_1+\nu_2+\cdots+\nu_m \text{ (length of the multi-index)},$$
$$\underline{\nu}!\equiv(\nu_1!)(\nu_2!)\cdots(\nu_m!),$$
$$w(\underline{\nu})\equiv|\underline{\nu}|!/\underline{\nu}! \text{ (weight of the multi-index)},$$
$$\underline{1}_\lambda\equiv(0,0,\ldots,0,1,0,\ldots,0) \quad (1 \text{ in the } \lambda \text{ th position}),$$
$$\underline{\nu}+\underline{\lambda}\equiv(\nu_1+\lambda_1,\nu_2+\lambda_2,\ldots,\nu_m+\lambda_m).$$

There are natural projections $\mathbf{J}^r\mathbf{Y}\to\mathbf{J}^s\mathbf{Y}$, for $r\geqslant s$; we recall that for any $r\geqslant1$, $\mathbf{J}^r\mathbf{Y}$ admits a natural structure of affine bundle over $\mathbf{J}^{r-1}\mathbf{Y}$.

(c) A Lagrangian of order r, that is a fibered morphism $L:\mathbf{J}^r\mathbf{Y}\to\Lambda^m\mathbf{T}^*\mathbf{X}$, which associates a volume m-form on $\mathbf{X}$ to any (local) section of $\mathbf{Y}$, together with its derivatives up to order r. The critical sections are those (local) sections $\sigma\in\Gamma(\mathbf{Y})$ which make stationary (in the sense of the calculus of variations) the action functionals, defined as the integrals $\int_{\mathbf{D}}L\circ j^r\sigma$, for any compact domain $\mathbf{D}\subset\mathbf{X}$. The Lagrangian L is locally expressed by a scalar density $L(x^\lambda,y^i_{\underline{\nu}})$, and can be equivalently represented by a global horizontal m-form $\Phi(L)=L(x^\lambda my^i_{\underline{\nu}})ds$ over $\mathbf{J}^r\mathbf{Y}$, where $ds=dx^1\wedge\cdots\wedge dx^m$.

In the sequel we shall also refer to the following objects, which, according to Ref. 26, are involved in the globalization of higher-order calculus of variations.

(d) The Poincaré–Cartan forms $\Theta(L,\Gamma)$, i.e., the global m-forms on $\mathbf{J}^{2r-1}\mathbf{Y}$, locally defined by the following expressions:

$$\Theta(L,\Gamma)_{|\mathbf{J}^{2r-1}\mathbf{U}} = \sum_{|\underline{\nu}|=0}^{r-1} f_i^{\underline{\nu},\lambda}(L,\Gamma)\omega_{\underline{\nu}}^i \wedge ds_\lambda + \Phi(L), \tag{2.1}$$

where Γ is a linear connection in the basis manifold $\mathbf{X}$; $ds_\lambda \equiv \partial_\lambda \rfloor ds$; $\Phi(L)$ is a shortcut for the local expression of the "pull-back" of the actual $\Phi(L)$ over $\mathbf{J}^{2r-1}\mathbf{Y}$; the one-forms $\omega_{\underline{\nu}}^i = dy_{\underline{\nu}}^i - y_{\underline{\nu}+1_\sigma}^i dx^\sigma$ belong to the natural basis for the contact one-forms on $\mathbf{J}^{2r-1}\mathbf{U}$; the coefficients $f_i^{\underline{\nu},\lambda}(L,\Gamma)$, called "components of the contact part of the Poincaré–Cartan form," satisfy global differential equations which locally read as follows:

$$f_i^{\underline{\nu},\lambda} = \frac{\partial L}{\partial y_{\underline{\nu}+1_\lambda}^i}, \quad \text{for } |\underline{\nu}| = r-1,$$
$$f_i^{\underline{\nu},\lambda} + d_\rho f_i^{\underline{\nu}+1_\lambda,\rho} = \frac{\partial L}{\partial y_{\underline{\nu}+1_\lambda}^i}, \quad \text{for } 0 \leqslant |\underline{\nu}| < r-1. \tag{2.2}$$

The existence and construction of global Poincaré–Cartan forms were discussed by various authors.[30–33] Poincaré–Cartan forms allow us to single out the critical sections of L by means of the following condition, which is equivalent to Euler–Lagrange equations: a section $\sigma \in \Gamma(\mathbf{Y})$ is a critical section iff

$$(j^{2r-1}\sigma)^*[i_\Xi d\Theta(L,\Gamma)] = 0, \tag{2.3}$$

for any vector field Ξ which is vertical with respect to the projection $\mathbf{J}^{2r-1}\mathbf{Y} \to \mathbf{X}$. The reader can easily check that in the case $m=1$, $r=1$, the definitions (2.1) and (2.2) lead to the well-known Poincaré–Cartan form of analytical mechanics: $\theta = p_i(dq^i - \dot{q}^i dt) + L\,dt$, $p_i = \partial L/\partial \dot{q}^i$, and that Eq. (2.3) reduces to the classical Euler–Lagrange equations $dp_i/dt - \partial L/\partial q^i = 0$.

(e) The "rth order Legendre bundle" $\mathbf{L}^r\mathbf{Y}$,[26] which is the vector bundle over $\mathbf{J}^{r-1}\mathbf{Y}$ defined by

$$\mathbf{L}^r\mathbf{Y} \equiv \mathbf{J}^{r-1}\mathbf{Y} \times_Y [\mathbf{V}^*(\mathbf{Y}) \otimes S_0^r(\mathbf{X}) \otimes \wedge^m \mathbf{T}^*\mathbf{X}], \tag{2.4}$$

where $\times_Y$ means the fibered product over $\mathbf{Y}$; $\mathbf{V}^*(\mathbf{Y})$ is the dual of the vertical tangent bundle of $\mathbf{Y}$; $\mathbf{S}_0^r(\mathbf{X})$ is the bundle of symmetric tensors of type $(r,0)$ over $\mathbf{X}$. A natural fibered chart on $\mathbf{L}^r\mathbf{Y}$ will be denoted by $(\mathbf{L}^r\mathbf{U}, x^\lambda, y_{\underline{\nu}}^i, p_i^{\underline{\mu}})$, $|\underline{\mu}| = r$, $|\underline{\nu}| \leqslant r-1$. The Legendre bundle is related to $\mathbf{J}^r\mathbf{Y}$ via the Legendre map Φ_L: $\mathbf{J}^r\mathbf{Y} \to \mathbf{L}^r\mathbf{Y}$, which is defined in the next section. According to Ref. 26, this bundle seems to be the most natural arena for the Hamiltonian description of dynamics, at least in the hyperregular case. In this sense, it plays a role in field theory which is analogous to the role played by the cotangent bundle $\mathbf{T}^*\mathbf{Q}$ in analytical mechanics. Further natural generalizations of the phase space of mechanics exist in higher-order field theory, namely the phase bundle and the momentum bundle, which in our opinion play different roles.[26] In particular, the phase bundle $\mathbf{P}(\mathbf{Y})$ is the vector bundle over $\mathbf{Y}$ defined by

$$\mathbf{P}(\mathbf{Y}) \equiv \mathbf{V}^*(\mathbf{Y}) \otimes \wedge^m \mathbf{T}^*\mathbf{X}. \tag{2.5}$$

With this definition the differential $d\Phi(L)$ of the Lagrangian form $\Phi(L)$ can be viewed as a section of $\mathbf{P}(\mathbf{J}^r\mathbf{Y})$ over $\mathbf{J}^r\mathbf{Y}$.

III. REGULARITY CONDITIONS AND LEGENDRE TRANSFORMATION

A. Survey about the hyperregular case

Here we summarize some fundamental ideas about the Hamiltonian formalism described in the previous papers,[25,26] which the reader may refer to for more details.

To any Lagrangian L of order r over $\mathbf{Y}$ we associated a bundle morphism $\Phi_L: \mathbf{J}^r\mathbf{Y} \to \mathbf{L}^r\mathbf{Y}$, over $\mathbf{J}^{r-1}\mathbf{Y}$, which in any natural fibered chart is expressed as

$$p_i^{\underline{\mu}} = \frac{\partial L}{\partial y_{\underline{\mu}}^i}, \quad |\underline{\mu}| = r. \tag{3.1}$$

The bundle morphism Φ_L is called the Legendre map. For reasons which shall be clear later, the image $\mathrm{Im}(\Phi_L) \subset \mathbf{L}^r\mathbf{Y}$ will be called the Hamiltonian constraint, which we denote by $\mathrm{HC}(L)$. Equations (3.1) can be suitably interpreted as the local equations defining $\mathrm{HC}(L)$. In general, to make a Hamiltonian description of dynamics possible, we should at least require that $\mathrm{HC}(L)$ be a fibered submanifold of $\mathbf{L}^r\mathbf{Y}$; this is probably the weakest requirement on L for this purpose.

Following Ref. (25), the regularity condition on Φ_L is locally expressed as

$$\det\left[\frac{\partial^2 L}{\partial y_{\underline{\mu}}^i \partial y_{\underline{\nu}}^j}\right] \neq 0, \quad |\underline{\mu}| = |\underline{\nu}| = r, \tag{3.2}$$

in any natural fibered chart. This is equivalent to the requirement that the Legendre map Φ_L be a local diffeomorphism. If Φ_L turns out to be a global diffeomorphism, we shall say that the Lagrangian L is hyperregular. The Hamiltonian constraint $\mathrm{HC}(L)$ coincides with the whole Legendre bundle if and only if L is hyperregular.

In the hyperregular case, the rth order jet prolongation of Φ_L transforms the ($2r$th order) Euler–Lagrange equation for L, which defines a fibered submanifold of $\mathbf{J}^{2r}\mathbf{Y}$, into an equivalent rth order equation over $\mathbf{L}^r\mathbf{Y}$, which defines a fibered submanifold of the bundle $\mathbf{J}^r(\mathbf{L}^r\mathbf{Y})$. Explicitly, we define the (global) Lagrangian Equation $\mathrm{LE}(L) \subset \mathbf{J}^{2r}\mathbf{Y}$ by the following local relations (in any natural chart):

$$\sum_{|\underline{\sigma}|=0}^{r} (-1)^{|\underline{\sigma}|} d_{\underline{\sigma}}\left(\frac{\partial L}{\partial y_{\underline{\sigma}}^i}\right) = 0. \tag{3.3}$$

The fibered submanifold $\mathrm{LE}(L)$ is globally well defined, since it is the inverse image of the zero section of the phase bundle $\mathbf{P}(\mathbf{J}^{2r}\mathbf{Y})$.

Now, we define a fibered submanifold $\mathrm{HE}(L) \subset \mathbf{J}^r(\mathbf{L}^r\mathbf{Y})$ by setting

$$\mathrm{HE}(L) \equiv (j^r\Phi_L \circ \iota^{r,r})\mathrm{LE}(L), \tag{3.4}$$

where $\iota^{r,r}$: $\mathbf{J}^{2r}\mathbf{Y} \to \mathbf{J}^r(\mathbf{J}^r\mathbf{Y})$ denotes the canonical embedding. In any natural fibered chart $(\mathbf{L}^r\mathbf{U}, x^\lambda, y_{\underline{\nu}}^i, p_i^{\underline{\mu}})$ of $\mathbf{L}^r\mathbf{Y}$, the fibered submanifold $\mathrm{HE}(L)$ is described by the following equations:

$$y^i_{\underline{\mu},\underline{\sigma}} - d_{\underline{\sigma}}\left[\frac{\partial H(L,U)}{\partial p^{\underline{\mu}}_{\underline{i}}}\right] = 0 \quad (|\underline{\sigma}| < r, |\underline{\mu}| = r),$$
$$\sum_{|\underline{\sigma}|=0}^{r-1} (-1)^{|\underline{\sigma}|} d_{\underline{\sigma}}\left[\frac{\partial H(L,U)}{\partial y^i_{\underline{\sigma}}}\right] - (-1)^r \sum_{|\underline{\mu}|=r} p^{\underline{\mu}}_{\underline{i},\underline{\mu}} = 0, \tag{3.5}$$

where a local m-form $\mathbf{H}(L,U) \equiv H(L,\mathrm{U})ds$, called local Hamiltonian, has been defined over $\mathbf{L}^r\mathbf{U}$ by the following prescription:

$$\mathbf{H}(L,\mathrm{U}) \equiv (\Phi_L)_* \left\{\left[\sum_{|\underline{\mu}|=r} y^i_{\underline{\mu}} \frac{\partial L}{\partial y^i_{\underline{\mu}}}\right] ds - \Phi(L)\right\}$$
$$\equiv \left\{\left[\left(\sum_{|\underline{\mu}|=r} y^i_{\underline{\mu}} \frac{\partial L}{\partial y^i_{\underline{\mu}}}\right) - L\right] \circ (\Phi_L)^{-1}\right\} ds. \tag{3.6}$$

Although the submanifold $\mathrm{HE}(L)$ is globally defined by (3.4), the local Hamiltonians defined in each chart by (3.6) cannot be patched together to define a global m-form. To overcome this difficulty, one can apply a globalization procedure based on the choice of a section of the affine fibration $\mathbf{J}^r\mathbf{Y} \rightarrow \mathbf{J}^{r-1}\mathbf{Y}$ (see Refs. 26 and 34) and show that there exists a (necessarily not unique) global form $\mathbf{H}$ which generates $\mathrm{HE}(L)$. Roughly speaking, this amounts to singling out a zero section of the affine bundle (in a suitably covariant way). Physically, that procedure corresponds to fixing a rest frame in particle mechanics.

For reasons of space and simplicity we cannot describe this method and we shall recall only how one can construct a global Hamiltonian $\mathbf{H}$ (we refer the reader to Refs. 26 and 34, where the method is worked out in full detail). Let $c: \mathbf{J}^{r-1}\mathbf{Y} \rightarrow \mathbf{J}^r\mathbf{Y}$ be a global section of the affine fibration $\mathbf{J}^r\mathbf{Y} \rightarrow \mathbf{J}^{r-1}\mathbf{Y}$, and let $j^{r-1}y \rightarrow (j^{r-1}y, c^i_{\underline{\mu}}(j^{r-1}y))$, with $|\underline{\mu}| = r$, be the local representation of the section c in a fibered chart. A global Hamiltonian $\mathbf{H}_c(L)$ can then be defined as follows:

$$\mathbf{H}_c(L)_{|L^rU}$$
$$\equiv (\Phi_L)_* \left\{\left[\sum_{|\underline{\mu}|=r} (y^i_{\underline{\mu}} - c^i_{\underline{\mu}}(j^{r-1}y)) \frac{\partial L}{\partial y^i_{\underline{\mu}}}\right] ds - \Phi(L)\right\}$$
$$\equiv \mathbf{H}(L,\mathrm{U}) - (\Phi_L)_* \left\{\left[\sum_{|\underline{\mu}|=r} c^i_{\underline{\mu}}(j^{r-1}y) \frac{\partial L}{\partial y^i_{\underline{\mu}}}\right] ds\right\}. \tag{3.7}$$

The corresponding Hamiltonian equations may be found in Ref. 26, Eq. (5.8).

When the global formalism is not strictly necessary, we shall restrict our further discussion to the local setting; accordingly, in the sequel we shall write $\mathbf{H}$ for $\mathbf{H}(L,\mathrm{U})$ and H for $H(L,\mathrm{U})$, omitting any explicit reference to the local character of this object.

Let us now remark that the strong regularity condition (3.2) plays an essential role in the above formalism, since the local Hamiltonian (3.6) is explicitly defined by a "push-forward" morphism over the Legendre map, which is defined only if Φ_L is a diffeomorphism. In order to extend this formalism to nonregular cases, one can hope to overcome the difficulty in two ways: either by proving, by an implicit function argument, that (local) m-forms $\mathbf{H}$ exist on the Hamiltonian constraint, such that the following holds,

$$(\Phi_L)^*[\mathbf{H}] = \left\{\left[\sum_{|\underline{\mu}|=r} y^i_{\underline{\mu}} \frac{\partial L}{\partial y^i_{\underline{\mu}}}\right] ds - \Phi(L)\right\}, \tag{3.8}$$

or, in an alternative but hopefully equivalent way, by restricting the Legendre map on suitably reduced spaces, in order to invert it explicitly. The former approach will be introduced in Ref. 28, where first-order problems are considered in their full generality; in the present paper we shall adopt the latter viewpoint, dealing with a restricted but physically interesting class of problems of arbitrary order.

B. Poincaré–Cartan forms over the Legendre bundle

Let us now re-express the relation between the Lagrangian and Hamiltonian picture in an alternative way, i.e., in terms of the equivalent description of dynamics generated by the Poincaré–Cartan forms. This will turn out to be useful in the sequel.

In this picture, as we said above, Euler–Lagrange equations assume the form (2.3). The Hamiltonian picture is obtained by taking the image of the Poincaré–Cartan forms $\Theta(L,\Gamma)$ under the Legendre map Φ_L. If the Lagrangian is hyperregular, one can set $\Theta(H,\Gamma) \equiv (\Phi_L)_*\Theta(L,\Gamma)$ and find the following explicit local expressions:

$$\Theta(H,\Gamma)_{|L^rU} = \left(\sum_{|\underline{\alpha}|=r-1} f^{\underline{\alpha},\lambda}_{\underline{i}}\, dy^i_{\underline{\alpha}} + \sum_{|\underline{\nu}|=0}^{r-2} f^{\underline{\nu},\lambda}_{\underline{i}}\, \omega^i_{\underline{\nu}}\right) \wedge ds_\lambda + \mathbf{H}, \tag{3.9}$$

where the coefficients $f^{\underline{\alpha},\lambda}_{\underline{i}}$ for $|\underline{\alpha}| = r-1$ are defined as follows:

$$f^{\underline{\alpha},\lambda}_{\underline{i}} = p^{\underline{\alpha}+1_\lambda}_{\underline{i}}, \quad \text{for } |\underline{\alpha}| = r-1, \tag{3.10}$$

while the remaining coefficients $f^{\underline{\nu},\lambda}_{\underline{i}}$, for $|\underline{\nu}| \leqslant r-2$, satisfy equations which are analogous to equations (2.2) above. The structural contact forms $\omega^i_{\underline{\nu}}$ are obtained via "push-forward" over Φ_L, namely by expressing the highest-order components $y^i_{\underline{\mu}}$, $|\underline{\mu}| = r$, as functions of the coordinates in $\mathbf{L}^r\mathbf{Y}$, via the inverse Legendre map:

$$\omega^i_{\underline{\nu}} = dy^i_{\underline{\nu}} - y^i_{\underline{\nu}+1_\lambda}(p^{\underline{\mu}}_{\underline{i}}, j^{r-1}y)dx^\lambda, \quad |\underline{\nu}| = r-1,$$
$$\omega^i_{\underline{\nu}} = dy^i_{\underline{\nu}} - y^i_{\underline{\nu}+1_\lambda}dx^\lambda, \quad |\underline{\nu}| \leqslant r-2. \tag{3.11}$$

Under these assumptions, the Hamiltonian equations (3.5) turn out to be equivalent to the following condition for Hamiltonian extremals $\rho \in \Gamma(\mathbf{L}^r\mathbf{Y})$:

$$(j^r\rho)^*[i_\Xi d\Theta(H,\Gamma)] = 0, \tag{3.12}$$

for any (vertical) vectorfield Ξ over $\mathbf{L}^r\mathbf{Y}$.

IV. LEGENDRE TRANSFORMATION IN NONHYPERREGULAR CASES

A. A typical nonhyperregular case

Let $L = L(j^ry)$ be a Lagrangian of order r on $\mathbf{Y}$. If the Lagrangian is not hyperregular, but the Legendre map has constant rank $\kappa < n$ $(n = h[(m+r-1)!/r!(m-1)!])$, there exist κ functionally independent functions $k^A(j^{r-1}y, y^i_{\underline{\mu}})$, $A = 1,\ldots,\kappa$, such that the functional dependence of L on the highest-order derivatives $y^i_{\underline{\mu}}$, $|\underline{\mu}| = r$, passes entirely through the functions k^A themselves. In other words, we can write

$$L(j^r y) = \tilde{L}(j^{r-1}y, k^A(j^{r-1}y, y^i_{\underline{\mu}})), \quad A = 1,\ldots,\kappa < n. \tag{4.1}$$

In this paper we shall fix our attention to the class of (nonhyperregular) Lagrangian field theories which satisfy the hypothesis (4.1), under the additional assumption that the functions k^A are affine combinations of the highest-order derivatives. In other words we assume that the following holds locally:

$$k^A(j^{r-1}y, y^i_{\underline{\mu}}) = \sum_{|\underline{\mu}|=r} \Lambda^{A\underline{\mu}}_i(j^{r-1}y) y^i_{\underline{\mu}} + T^A(j^{r-1}y), \tag{4.2}$$

where Λ is an $n \times \kappa$ matrix of rank κ. This situation, which could seem to be rather exceptional from a purely mathematical viewpoint, is in fact very often encountered in the physical literature; for instance, both general relativity (as well as all "alternative" metric theories of gravitation) and Yang–Mills theory display this feature.

Our aim is to show that under a suitable regularity condition (which is weaker than hyperregularity) one can obtain for the class of Lagrangians considered an equivalent Hamiltonian description of dynamics. This description, which is close to the ordinary one for the hyperregular case, is based on a method which allows to find a natural parametrization of the Hamiltonian constraint HC(L) and, consequently, to define a Hamiltonian directly on the Hamiltonian constraint itself.

Let us first restate our hypotheses (4.1) and (4.2) in a more intrinsic way. Let $\mathbf{K}$ be a vector bundle, of rank κ, over $\mathbf{J}^{r-1}\mathbf{Y}$, and let $(x^\lambda, y^i_{\underline{\nu}}, k^A)$, $A = 1,\ldots,\kappa$, $|\underline{\nu}| \leqslant r-1$, be a (local) fibered parametrization for $\mathbf{K}$. In the sequel, for the sake of brevity, we shall sometimes write $(j^{r-1}y, k)$ instead of $(x^\lambda, y^i_{\underline{\nu}}, k^A)$. Let also $s_{\mathbf{K}}$: $\mathbf{J}^r\mathbf{Y} \to \mathbf{K}$ be a surjective morphism of affine bundles over $\mathbf{J}^{r-1}\mathbf{Y}$, i.e., a morphism locally defined by relations of the form (4.2). Our hypothesis on the Lagrangian can be thence restated as follows: there exists a reduced Lagrangian $\tilde{L}$: $\mathbf{K} \to \Lambda^m\mathbf{T}^*\mathbf{X}$ such that the original Lagrangian L factors as

$$L = \tilde{L} \circ s_{\mathbf{K}}. \tag{4.3}$$

As usual, $\tilde{L}$ can be represented by a horizontal m-form $\Phi(\tilde{L}) = \tilde{L}(j^{r-1}y, \kappa^A)\, ds$.

The differential $d\Phi(\tilde{L})$ of the reduced Lagrangian form $\Phi(\tilde{L})$ can be viewed as a section of the phase bundle $\mathbf{P}(\mathbf{K})$ over $\mathbf{K}$. Since $\mathbf{K}$ is a vector bundle over $\mathbf{J}^{r-1}\mathbf{Y}$, there is an isomorphism

$$\mathbf{P}(\mathbf{K}) \cong \mathbf{P}(\mathbf{J}^{r-1}\mathbf{Y}) \underset{J^{r-1}Y}{\times} \mathbf{K} \underset{J^{r-1}Y}{\times} [\mathbf{K}^* \otimes \Lambda^m\mathbf{T}^*\mathbf{X}], \tag{4.4}$$

where $\mathbf{K}^*$ is the dual vector bundle of $\mathbf{K} \to \mathbf{J}^{r-1}\mathbf{Y}$.

Using a natural projection we can define a map

$$\partial\tilde{L}: \mathbf{K} \to \mathbf{K}^* \otimes \Lambda^m\mathbf{T}^*\mathbf{X}, \tag{4.5}$$

which is a morphism of bundles over $\mathbf{J}^{r-1}\mathbf{Y}$ and which represents the differential of $\Phi(\tilde{L})$ with respect to the variables k^A alone. (It is in fact the vertical differential of $\tilde{L}$ with respect to the natural projection $\mathbf{K} \to \mathbf{J}^{r-1}\mathbf{Y}$.) The local representation of (4.5) is the following:

$$\pi_A = \frac{\partial\tilde{L}}{\partial k^A}, \tag{4.6}$$

where $(j^{r-1}y, \pi_A)$ is a natural coordinate chart in $\mathbf{K}^*$.

We say that the original Lagrangian L is $\mathbf{K}$-hyperregular iff the reduced Lagrangian $\tilde{L}$ is hyperregular, i.e., the morphism $\partial\tilde{L}$ is a bundle isomorphism between $\mathbf{K}$ and $\mathbf{K}^* \otimes \Lambda^m\mathbf{T}^*\mathbf{X}$. This implies that *locally* $\tilde{L}$ satisfies the condition:

$$\det\left[\frac{\partial^2\tilde{L}}{\partial k^A \partial k^B}\right] \neq 0, \tag{4.7}$$

which can be called the condition for (*local*) $\mathbf{K}$-regularity for L.

Let us remark now that it is not difficult to define a dual morphism:

$$(s_{\mathbf{K}})^*: \mathbf{K}^* \otimes \Lambda^m\mathbf{T}^*\mathbf{X} \to \mathbf{L}^r\mathbf{Y}, \tag{4.8}$$

by using the fact that $\mathbf{K} \to \mathbf{J}^{r-1}\mathbf{Y}$ is a vector bundle and $s_{\mathbf{K}}$: $\mathbf{J}^r\mathbf{Y} \to \mathbf{K}$ is an affine morphism over $\mathbf{J}^{r-1}\mathbf{Y}$. The morphism $(s_{\mathbf{K}})^*$ is locally represented by

$$p^{\underline{\mu}}_i = \pi_A \left(\frac{\partial k^A}{\partial y^i_{\underline{\mu}}}\right) = \pi_A \Lambda^{A\underline{\mu}}_i(j^{r-1}y), \tag{4.9}$$

where the last equality follows from (4.2).

Consider now the following diagram:

$$\begin{array}{ccc} \mathbf{K} & \xrightarrow{\ \partial\tilde{L}\ } & \mathbf{K}^* \otimes \Lambda^m\mathbf{T}^*\mathbf{X} \\ {\scriptstyle s_{\mathbf{K}}}\uparrow & & \downarrow{\scriptstyle (s_{\mathbf{K}})^*} \\ \mathbf{J}^r\mathbf{Y} & \xrightarrow[\ \Phi_L\ ]{} & \mathbf{L}^r\mathbf{Y} \end{array}$$

This diagram is commutative, i.e., we have

$$\Phi_L = (s_{\mathbf{K}})^* \circ \partial\tilde{L} \circ s_{\mathbf{K}}. \tag{4.10}$$

Relation (4.10) is in fact the global version of the chain rule

$$\frac{\partial L}{\partial y^i_{\underline{\mu}}} = \frac{\partial\tilde{L}}{\partial k^A} \frac{\partial k^A}{\partial y^i_{\underline{\mu}}}, \quad |\underline{\mu}| = r. \tag{4.11}$$

Since the Hamiltonian constraint $\mathrm{HC}(L) \equiv \mathrm{Im}(\Phi_L) \subseteq \mathbf{L}^r\mathbf{Y}$ is represented by the local equations (3.1), we see from (4.9) and (4.11) that whenever $\partial\tilde{L}$ is a surjective morphism the following holds:

$$\mathrm{HC}(L) \equiv \mathrm{Im}(\Phi_L) = \mathrm{Im}[(s_K)^*] \subseteq \mathbf{L}^r\mathbf{Y}. \tag{4.12}$$

Since $(s_{\mathbf{K}})^*$ is an injection, the equality (4.12) assures that the bundles $\mathbf{K}^* \otimes \Lambda^m\mathbf{T}^*\mathbf{X}$ and $\mathrm{HC}(L)$ are isomorphic. Therefore, $(j^{r-1}y, \pi)$ can be considered, without any confusion, as coordinates on the Hamiltonian constraint HC(L).

From now on we shall assume that the Lagrangian L is $\mathbf{K}$-hyperregular [for local purposes it would be enough to require that the local condition (4.7) holds]. Under this hypothesis the relation (4.12) holds *a fortiori*. We can thence define by range-restriction a reduced Legendre map $\Phi_{\tilde{L}}: \mathbf{K} \to \mathrm{HC}(L)$ by setting

$$\Phi_L \equiv (s_K)^* \circ \Phi_L, \tag{4.13}$$

or equivalently:

$$\Phi_L = \Phi_{\tilde{L}} \circ s_K. \tag{4.14}$$

This is of course a bundle isomorphism, which is locally expressed by a relation which is formally identical to (4.6), provided the coordinates π_A are now considered as coordinates in HC(L).

This allows a bijective Legendre transformation from the truly dynamical part **K** of the velocity space $\mathbf{J}^r\mathbf{Y}$ onto the Hamiltonian constraint HC(L) in the phase space $\mathbf{L}^r\mathbf{Y}$. In fact, we can define the reduced Hamiltonian $\widetilde{H}$: HC(L) $\to \Lambda^m\mathbf{T}^*\mathbf{X}$ by setting locally

$$\widetilde{H} = \left[\frac{\partial \widetilde{L}}{\partial k^A} k^A - \widetilde{L}\right]\circ(\Phi_{\widetilde{L}})^{-1}$$
$$= \pi_A\left[k^A\circ(\Phi_{\widetilde{L}})^{-1}\right] - \widetilde{L}\circ(\Phi_{\widetilde{L}})^{-1}$$
$$= \pi_A k^A(j^{r-1}y,\pi) - \widetilde{L}(j^{r-1}y,\pi), \tag{4.15}$$

where $\widetilde{L}$ and k^A appear as functions on HC(L) via the inverse Legendre map $(\Phi_{\widetilde{L}})^{-1}$. We stress that the local expressions (4.16) define in fact a global m-form on HC(L). It is easily checked that in any fibered chart **U** the following holds:

$$H(L,\mathbf{U})\circ\Phi_L \equiv \sum_{|\underline{\mu}| = r-1} y^i_{\underline{\mu}} \frac{\partial L}{\partial y^i_{\underline{\mu}}} - L = (\widetilde{H} - \pi_A T^A)\circ\Phi_L, \tag{4.16}$$

which in a suitable sense defines in HC(L) a local m-form **H** satisfying (3.8).

Therefore, any Poincaré–Cartan form $\Theta(L,\Gamma)$ turns out to be the pull-back onto $\mathbf{J}^{2r-1}\mathbf{Y}$ of the global m-form $\Theta(\widetilde{H},\Gamma)$ defined over $\mathbf{J}^{r-1}$HC(L) by the following local equations:

$$\Theta(\widetilde{H},\Gamma) = \left[\sum_{|\underline{\mu}| = r} w(\underline{\mu})\pi_A \Lambda^{A\underline{\mu}}_i\, dy^i_{\underline{\mu}} + \sum_{|\underline{\nu}| = 0}^{r-2} f^{\underline{\nu}\lambda}_i \omega^i_{\underline{\nu}}\right]\wedge ds_\lambda - (H - \pi_A T^A) ds. \tag{4.17}$$

Here, the coefficients $f^{\underline{\nu}\lambda}_i$, $|\underline{\nu}| \leqslant r-2$, as in Eq. (3.9), are expressed in terms of the natural coordinates in $\mathbf{J}^{r-1}$HC(L) by means of the Legendre map.

Moreover, the appropriate Hamiltonian equations can be obtained by specifying that the Hamiltonian extremals are those local sections ρ: $\mathbf{W}\to$HC(L) which satisfy the equations

$$(j^r\rho)^*[i_\Xi\, d\Theta(\widetilde{H},\Gamma)] = 0, \tag{4.18}$$

for any (vertical) vectorfield Ξ over $\mathbf{J}^{r-1}$ HC(L). A variational principle can be associated to the Hamiltonian equations by means of the Helmholtz Lagrangian $L_{\widetilde{H}}$, defined by

$$L_{\widetilde{H}}(j^r y,\pi) \equiv \pi_A k^A(j^r y) - \widetilde{H}. \tag{4.19}$$

We shall consider this expression again in the remarks below, in connection with the concept of dual Lagrangians.

B. Remarks

(a) The hypotheses (4.6) and (4.7) are both essential in our discussion, but it is worthwhile to consider separately to what extent each one of them restricts the domain of application of our method. If the dependence of the functions k^A on the highest-order derivatives of the fields y^i is nonlinear, one has no direct way to define a morphism $(s_K)^*$ which closes diagram (1). Let us stress, however, that this occurrence does not imply the absence of a Hamiltonian description in a more general sense[28]. The existence of the Hamilton function is in fact assured under weaker conditions, as it can be easily seen from worked examples. On the other hand, the Hessian of $\widetilde{L}$ with respect to k^A may be degenerate, but still of constant rank: in this case, HC(L) is still embedded as a subbundle into $\mathbf{L}^r\mathbf{Y}$ (its rank, i.e., fiber dimension, is obviously the rank of the Hessian of $\widetilde{L}$), but the κ functions π_A are not independent and they do not provide a parametrization of HC(L). Nevertheless, the correct parametrization for this case could possibly be obtained by a suitable choice of a maximal independent set among the functions π_A. Accordingly, the expressions (4.6) and (4.19) would still be valid, provided the remaining functions π_A are assumed to be functionally dependent on the former ones and provided this is taken into account when varying $L_{\widetilde{H}}$.

(b) Allowing the rank κ to be equal to n in the above framework one reobtains, in a different parametrization, the results already known for the hyperregular case. We also remark that in the particular case $r = 1$ the regularity condition (3.15) turns out to be automatically satisfied, so that our results for $r = 1$ are in agreement with those presented in Ref. 18 for the general situation.

(c) Let L and L' be two Lagrangians of order r and s, respectively, defined on two different fibered manifolds **Y** and **Y**$'$ over **X**. We say that L and L' are dynamically equivalent if a fibered morphism ϕ: $\mathbf{J}^r\mathbf{Y}\to\mathbf{J}^s\mathbf{Y}'$ exists such that the following holds: a section $\sigma\in\Gamma(\mathbf{Y})$ is a Lagrangian extremal for L if and only if the image section $\rho = \phi(j^r\sigma)\in\Gamma(\mathbf{J}^s\mathbf{Y}')$ is the sth order prolongation of a Lagrangian extremal for L'. It is intuitively clear, and it could be shown explicitly, that the Legendre transformation establishes a dynamical equivalence between the original rth order variational principle on **Y** and a variational principle of the same order r on the vector bundle $\mathbf{L}^r\mathbf{Y}\equiv\mathbf{V}^*(\mathbf{Y})\otimes\mathbf{S}^r_0(\mathbf{X})\otimes\Lambda^m\mathbf{T}^*\mathbf{X}$, which is parametrized by the coordinates $(x^\lambda,y^i,p^{\underline{\mu}}_i)$. This variational principle is described by the Helmholtz Lagrangian $L_{\widetilde{H}}$, which depends linearly on the derivatives of order r in such a way to generate equations of order $s\equiv\inf\{2(r-1),1\}$. It happens sometimes (as we shall see in some of the examples below, namely the affine gravitational theories and the Weyl's conformal gauge theory), that a further step makes possible to eliminate completely the dependence of $L_{\widetilde{H}}$ on the original dynamical variables y^i. In this case, one finds a dual variational principle which is formulated in terms of the momenta $p^{\underline{\sigma}}_i$ and their jet prolongations only. In the known examples this occurrences seems to be purely accidental, but in fact it is deeply connected, in general, with the geometric structure of the bundle $\mathbf{L}^r\mathbf{Y}$ and with its possible splittings. This subject is currently under investigation. Whenever this third equivalent representation of dynamics can be obtained, the connections between the corresponding three sets of equations turn out to provide a particular example of a Lie–Bäcklund transformation, which is a subject of considerable interest in the formal theory of PDEs (for a thorough presentation of this subject, see Ref. 20).

(d) It should be clear at this point that the physical motivation for the Hamiltonian formulation of field theories, in the sense presented hereby, is completely different from the usual motivation for the other possible "Hamiltonian pictures" already mentioned in the Introduction. In fact, these latter ones are considered in order to provide a time evolution description of field dynamics. Our approach, on the contrary, is aimed to a deeper understanding of the structure of physical theories. In particular, we hope that investigating alternative formulations of the same theory, through different sets of dynamical variables, can help to emphasize the distinction between the physical contents and the mathematical machinery in field-theoretical models.

V. EXAMPLES

A. First-order affine gravitational theories and the "Einstein–Eddington prescription"

The so-called affine theories of gravitation are a natural field of application of the methods presented above. In fact, it is well known that the requirement of general covariance implies that a Lagrangian can depend on the first derivatives of a linear connection only through the Riemann tensor and the covariant derivative of the torsion tensor. Thus, when a linear connection is assumed to be the dynamical variable, one is led directly to the situation described by (4.5) and (4.6). When dealing with a symmetric connection (or more generally with a Lagrangian independent of the torsion), the functions k^A are thus to be identified either with the independent components of the Riemann tensor or with some suitable combinations of these ones. The correct identification of the functions k^A is usually determined by the regularity condition (4.7): the choice could be not unique, but different choices compatible with (4.7) lead to different parametrizations of the same bundle **K**.

As a first example, let us recall the affine formulation of general relativity, due to Eddington and Einstein.[1–3] According to the definitions adopted in this paper, we describe this model by assuming **X** to be a four-manifold representing the physical space-time, **Y** to be the bundle of linear symmetric connections over **X**, and the Lagrangian to be defined by setting

$$\Phi(L_E) = z\sqrt{|\det(G_{\alpha\beta})|}\,ds, \tag{5.1}$$

where $G_{\alpha\beta} \equiv \Gamma^{\lambda}_{\alpha\beta,\lambda} - \Gamma^{\lambda}_{\lambda(\alpha,\beta)} + \Gamma^{\lambda}_{\alpha\beta}\Gamma^{\sigma}_{\lambda\sigma} - \Gamma^{\lambda}_{\alpha\sigma}\Gamma^{\sigma}_{\beta\lambda}$ is the symmetric part of the Ricci tensor associated with a section Γ of **Y**, and $z \neq 0$ is a real number. Setting for brevity

$$J^{\lambda\mu\nu}_{\sigma} = -(z/2)(G^{\lambda\mu}\delta^{\nu}_{\sigma} - G^{\nu/\lambda}\delta^{\mu}_{\sigma})\sqrt{|\det(G_{\alpha\beta})|}, \tag{5.2}$$

where $G^{\lambda\mu}$ is the inverse of $G_{\lambda\mu}$, i.e., $G^{\lambda\mu}G_{\mu\nu} = \delta^{\lambda}_{\nu}$, we obtain the Lagrange equations for L_E in the following form:

$$\nabla_{\alpha}J^{\lambda\mu\alpha}_{\nu} \equiv \partial_{\alpha}J^{\lambda\mu\alpha}_{\nu} + \Gamma^{\lambda}_{\sigma\alpha}J^{\sigma\mu\alpha}_{\nu} + \Gamma^{\mu}_{\sigma\alpha}J^{\lambda\sigma\alpha}_{\nu} - \Gamma^{\sigma}_{\nu\alpha}J^{\lambda\mu\alpha}_{\sigma} = 0. \tag{5.3}$$

To apply our procedure, we assume that **K** is the quotient bundle $\mathbf{J}^1\mathbf{Y}/\approx$, under the equivalence relation $j^1\Gamma \approx j^1\Gamma'$ iff $G_{\lambda\mu}(j^1\Gamma) = G_{\lambda\mu}(j^1\Gamma')$. The fiber of **K** is thence spanned by the ten independent components of $G_{\lambda\mu}$. It is easy to check that the following holds:

$$\det\left|\left|\frac{\partial^2 L_E}{\partial G_{\alpha\beta}\partial G_{\lambda\mu}}\right|\right| \neq 0. \tag{5.4}$$

We now set

$$\pi^{\alpha\beta} \equiv \frac{\partial L_E}{\partial G_{\alpha\beta}} = \frac{z}{2}G^{\alpha\beta}\sqrt{|\det(G_{\lambda\mu})|}. \tag{5.5}$$

It is convenient, in this case, to replace the contravariant symmetric density $\pi^{\alpha\beta}$ with the associated metric tensor $\gamma^{\alpha\beta}$ defined by

$$\gamma^{\alpha\beta} \equiv \frac{\pi^{\alpha\beta}}{\sqrt{|\det(\pi^{\mu\nu})|}} = \frac{2}{z}G^{\alpha\beta}. \tag{5.6}$$

We thus find

$$\tilde{H} = (4/z)\sqrt{|\det(\gamma_{\mu\nu})|}, \tag{5.7}$$

$$L_{\tilde{H}}(\gamma, j^1\Gamma) = [\gamma^{\alpha\beta}G_{\alpha\beta}(j^1\Gamma) - 4/z]\sqrt{|\det(\gamma_{\mu\nu})|}. \tag{5.8}$$

$L_{\tilde{H}}$ is a metric-affine Lagrangian, which is very close to the standard Einstein–Hilbert Lagrangian for the purely metric version of general relativity with a cosmological constant. In fact, $L_{\tilde{H}}$ differs from that one only because $G_{\alpha\beta}$ is the Ricci tensor associated to the dynamical connection $\Gamma^{\lambda}_{\alpha\beta}$ rather than the Ricci tensor $R_{\alpha\beta}$ associated to the Levi-Civita connection $\{^{\lambda}_{\alpha\beta}\}$ of the metric $\gamma_{\alpha\beta}$. However, if one uses the definition (5.6) and replaces $G_{\alpha\beta}$ with $(z/2)\,\gamma_{\alpha\beta}$ in the relation (5.2), the dynamical equations (5.3) become

$$\Gamma^{\lambda}{}_{\alpha\beta} = \{^{\lambda}_{\alpha\beta}\}. \tag{5.9}$$

We can thus substitute $R_{\alpha\beta}(j^2\gamma)$ for $G_{\alpha\beta}(j^2\Gamma)$ in (5.8), and show in this way a complete dynamical equivalence between general relativity and the affine theory based on the Lagrangian (5.1). The definition (5.6), which introduces in a canonical way a metric tensorfield in the framework of a purely affine theory, coincides with the well-known "Eddington–Einstein prescription" (shortly "E–E prescription"). We stress that the procedure to obtain the Lagrangian $L_{\tilde{H}}$ follows exactly the general method described in the previous section. The further step, which consists in replacing $G_{\alpha\beta}(j^1\Gamma)$ with $R_{\alpha\beta}(j^2\gamma)$, is instead based on the particular geometric features of the objects involved.

This procedure has been generalized to any affine Lagrangian of the type

$$L = L(G_{\alpha\beta}, \Gamma^{\lambda}{}_{\alpha\beta}). \tag{5.10}$$

The application of the E–E prescription

$$\gamma^{\alpha\beta} \equiv \frac{\partial L}{\partial G_{\alpha\beta}}\frac{1}{\sqrt{\det\|\partial L/\partial G_{\mu\nu}\|}}, \tag{5.11}$$

and the consequent metric reformulation of the theory, have led to the general result that any such theory is dynamically equivalent to general relativity.[5] As a consequence, all these theories can be considered as affine formulations of general relativity. Moreover, starting from a generic connection with torsion, it has been shown in recent years by Kijowski and one of us (M. F.) that a whole family of unified Lagrangians exists, such that the corresponding metric (i.e., Hamiltonian) counterpart generates explicitly the coupled Einstein–Maxwell dynamical equations.[6,7]

Let us remark once more that the identification of the E–E prescription with a Legendre transformation has been suggested several times in the previous literature. The main motivation of this paper was in fact to provide a satisfactory geometric motivation for this identification.

B. Yang–Mills theory

The Yang–Mills gauge model provides another example of a possible application of our formalism. When the gauge field is the only field present, the configuration bundle **Y** can be identified with the bundle **C** of all connections of a principal bundle **P** over space-time **X**, with structure group **G**. The reduced Lagrangian $\tilde{L}_{\rm YM}$ depends on the curvature form of a connection **A**, i.e., on the coefficients

$$F^a{}_{\mu\nu} = \partial_\mu A^a{}_\nu - \partial_\nu A^a{}_\nu + c^a{}_{bc} A^b{}_\mu A^c{}_\nu, \qquad (5.12)$$

where c^a_{bc} are the structure constants of the Lie algebra g; it is defined by setting

$$\tilde{L}_{\rm YM} = \tfrac{1}{4} F^a{}_{\mu\nu} F^b{}_{\rho\sigma} g^{\mu\rho} g^{\nu\sigma} \eta_{ab} \sqrt{|\det(g_{\alpha\beta})|}, \qquad (5.13)$$

whereby $g^{\alpha\beta}$ is a metric on **X**, fixed *a priori*, and η_{ab} is an invariant metric on **g**. One can immediately see that $\tilde{L}_{\rm YM}$ depends in a regular way on the components $F^a_{\mu\nu}$, and one easily finds

$$\pi_a{}^{\mu\nu} \equiv 2 \frac{\partial \tilde{L}_{\rm YM}}{\partial F^a_{\mu\nu}} = F^b{}_{\rho\sigma} g^{\mu\rho} g^{\nu\sigma} \eta_{ab} \sqrt{|\det(g_{\alpha\beta})|}, \qquad (5.14)$$

$$\tilde{H} = \frac{1}{4\sqrt{|\det(g_{\alpha\beta})|}} \pi_a{}^{\mu\nu} \pi_b{}^{\rho\sigma} g_{\mu\rho} g_{\nu\sigma} \eta^{ab}, \qquad (5.15)$$

$$H = \frac{1}{4\sqrt{|\det(g_{\alpha\beta})|}} \pi_a{}^{\mu\nu} \pi_b{}^{\rho\sigma} g_{\mu\rho} g_{\nu\sigma} \eta^{ab} - \tfrac{1}{2} \pi_a{}^{\mu\nu} c^a{}_{bc} A^b{}_\mu A^c{}_\nu, \qquad (5.16)$$

$$L_{\tilde{H}} = \tfrac{1}{2} \pi_a{}^{\mu\nu} F^a{}_{\mu\nu} - \frac{1}{4\sqrt{|\det(g_{\alpha\beta})|}} \pi_a{}^{\mu\nu} \pi_b{}^{\rho\sigma} g_{\mu\rho} g_{\nu\sigma} \eta^{ab}. \qquad (5.17)$$

From the transformation laws of $F^a{}_{\mu\nu}$ and of $\pi_a{}^{\mu\nu}$ as given by (5.14) one easily realizes that the appropriate reduced bundle and its dual for a free Yang–Mills theory are

$$\mathbf{K}_{\rm YM} = \mathbf{g} \otimes \Lambda^2 \mathbf{T}^*\mathbf{X}, \quad \mathbf{K}^*_{\rm YM} = \mathbf{g}^* \otimes \Lambda^{m-2} \mathbf{T}^*\mathbf{X}.$$

The quadratic Lagrangian $\tilde{L}_{\rm YM}$ is a field-theoretic analog, from our viewpoint, of the Lagrangian $L = \frac{1}{2} m\dot{q}^2$ of particle mechanics. The Legendre transformation is straightforward and does not seem to provide any really new information for this case.

On the other hand, let us shortly comment on a more general Lagrangian coupling the gauge field with a charged field, i.e.,

$$L = L_{\rm YM}(j^1 A) + L_\Psi(A, j^1\Psi), \qquad (5.18)$$

where the field Ψ is a section of a suitable vector bundle **E** associated to the principal bundle **P**. The configuration bundle is now $\mathbf{Y} \equiv \mathbf{C} \times_X \mathbf{E}$. Following the steps described above for the Lagrangian $L_{\rm YM}$, we can perform a partial transformation

$$\mathbf{J}^1(\mathbf{C} \times_X \mathbf{E}) \equiv \mathbf{J}^1\mathbf{C} \times_X \mathbf{J}^1\mathbf{E} \to \mathbf{L}^1\mathbf{C} \times_X \mathbf{J}^1\mathbf{E},$$

whereby H plays the role of a "Routh function" rather than the role of a Hamiltonian. In this case the relevant reduced bundles will be the following:

$$\mathbf{K} = [\mathbf{K}_{\rm YM} \oplus \mathbf{J}^1\mathbf{E}] \times_X \mathbf{C}, \quad \mathbf{K}^* = [\mathbf{K}^*_{\rm YM} \oplus \mathbf{J}^1\mathbf{E}] \times_X \mathbf{C}.$$

If the interaction Lagrangian L_Ψ is sufficiently regular with respect to Ψ one could also perform a total Legendre transformation $\mathbf{J}^1(\mathbf{C} \times_X \mathbf{E}) \to \mathbf{L}^1(\mathbf{C} \times_X \mathbf{E})$. In this case, a partial transformation $\mathbf{J}^1\mathbf{C} \times_X \mathbf{J}^1\mathbf{E} \to \mathbf{J}^1\mathbf{C} \times_X \mathbf{L}^1\mathbf{E}$ is also possible. We cannot exclude that a suitable combination of these transformations may reveal some relationship between different theories of some physical interest. This problem is currently under investigation and will form the subject of further papers.

C. Second-order metric gravitational theories

We recall the results presented in our previous paper,[10] to which we refer the reader for further details.

A purely metric gravitational theory is a second-order theory on the bundle of all Lorentz metrics $\gamma_{\alpha\beta}$ over a four-manifold **X**; general covariance implies that the Lagrangian depends on the first and second derivatives of $\gamma_{\alpha\beta}$ only through the Riemann tensor. For simplicity, we shall deal with Lagrangians depending only on the Ricci tensor $K_{\alpha\beta}(j^2\gamma)$ of $\gamma_{\alpha\beta}$, defined as usual by

$$K_{\alpha\beta} \equiv \{^{\lambda}_{\alpha\beta}\}_{,\lambda} - \{^{\lambda}_{\lambda\alpha}\}_{,\beta} + \{^{\lambda}_{\alpha\beta}\}\{^{\sigma}_{\lambda\sigma}\} - \{^{\lambda}_{\alpha\sigma}\}\{^{\sigma}_{\beta\lambda}\};$$

accordingly, we set L to be

$$L(j^2\gamma) = \tilde{L}[K_{\alpha\beta}(j^2\gamma), \gamma_{\alpha\beta}]. \qquad (5.19)$$

The Euler–Lagrange equations are in general of the fourth order:

$$j^{\alpha\beta} + \tfrac{1}{2}(\pi^{\beta\nu}{}_{;\mu\nu}\gamma^{\alpha\mu} + \pi^{\alpha\nu}{}_{;\mu\nu}\gamma^{\beta\mu} - \pi^{\mu\nu}{}_{;\mu\nu}\gamma^{\alpha\beta} - \pi^{\alpha\beta}{}_{;\mu\nu}\gamma^{\mu\nu}) = 0, \qquad (5.20)$$

where we have set

$$j^{\alpha\beta} = \frac{\partial \tilde{L}}{\partial \gamma_{\alpha\beta}}, \qquad (5.21)$$

$$\pi^{\alpha\beta} = \frac{\partial \tilde{L}}{\partial K_{\alpha\beta}}. \qquad (5.22)$$

The regularity condition then becomes the following:

$$\det \left\| \frac{\partial^2 \tilde{L}}{\partial K_{\alpha\beta}\, \partial K_{\lambda\mu}} \right\| \neq 0. \qquad (5.23)$$

As in the case of affine theories, it turns out to be convenient to express the Hamiltonian formulation through the metric tensor $g^{\alpha\beta}$ associated to the momentum $\pi^{\alpha\beta}$, i.e., to follow the E–E prescription, which in this case reads as follows:

$$g^{\alpha\beta} \equiv \frac{\partial \tilde{L}}{\partial K_{\alpha\beta}} \frac{1}{\sqrt{|\det(\partial \tilde{L}/\partial K_{\mu\nu})|}}. \qquad (5.24)$$

Following our general rules we obtain

$$L_{\tilde{H}} = g^{\alpha\beta} K_{\alpha\beta}(j^2\gamma)\sqrt{|\det(g_{\mu\nu})|} - \tilde{H}(g^{\alpha\beta}, \gamma_{\mu\nu}), \qquad (5.25)$$

where $\tilde{H}(g^{\alpha\beta}, \gamma_{\mu\nu})$ is easy to calculate (its expression can be found in Ref. 10). We introduce now the tensor $Q^{\lambda}_{\alpha\beta}(j^1g, j^1\gamma)$ by setting

$$Q^{\lambda}_{\alpha\beta} \equiv \tfrac{1}{2}\gamma^{\lambda\sigma}(\nabla_\beta \gamma_{\alpha\sigma} + \nabla_\alpha \gamma_{\sigma\beta} - \nabla_\sigma \gamma_{\alpha\beta}), \tag{5.26}$$

where ∇_α denotes the covariant derivative with respect to the Levi-Civita connection of the metric $g_{\alpha\beta}$. This allows us to re-express $K_{\alpha\beta}(j^2\gamma)$ through the Ricci tensor $R_{\alpha\beta}(j^2g)$ of the new metric $g_{\alpha\beta}$, so that the following relation is easily inferred:

$$L_{\widetilde{H}} = [g^{\alpha\beta}R_{\alpha\beta}(j^2g) + g^{\alpha\beta}(Q^{\lambda}_{\alpha\beta}Q^{\sigma}_{\sigma\lambda} - Q^{\lambda}_{\sigma\beta}Q^{\sigma}_{\alpha\lambda})]\sqrt{|\det(g_{\mu\nu})|} - \widetilde{H}(g^{\alpha\beta},\gamma_{\mu\nu}) \tag{5.27}$$

(up to a divergence term). In the expression (5.27) an Einstein–Hilbert term for the new metric $g_{\alpha\beta}$ appears, while the original metric $\gamma_{\alpha\beta}$ appears together with its first derivatives only, and it is coupled to $g_{\alpha\beta}$ as if it were an external matter field in general relativity.

If external matterfields are present in the original Lagrangian, a partial Legendre transformation can be performed without any substantial difference, as it was shown in Ref. 10. If the derivatives of the external fields appear only up to the first order, as it commonly happens, this "partial" transformation turns out to be the total one. Even if the Legendre transformation does not seem to affect the external fields, however, their coupling with the field $\gamma_{\alpha\beta}$ in the Hamiltonian picture is obviously different from the original one. The interaction of $\gamma_{\alpha\beta}$, $g_{\alpha\beta}$ and the external fields after the transformation depends in a nontrivial way on the structure of the original Lagrangian. This occurrence is physically relevant: in particular, in fact, the transformation does not preserve the condition of minimal coupling; conversely, it is possible to transform a theory with nonminimal interaction into a minimally coupled one.

A different situation occurs when the Lagrangian depends only on the scalar curvature $K(j^2\gamma) \equiv \gamma^{\alpha\beta}K_{\alpha\beta}(j^2\gamma)$:

$$L(j^2\gamma) = f[K(j^2\gamma)]\sqrt{|\det(\gamma_{\mu\nu})|}, \tag{5.28}$$

where f satisfies the regularity condition

$$\frac{d^2f}{dK^2} = 0. \tag{5.29}$$

In this case **K** is a bundle of rank one, fiberwise spanned by the variable

$$\pi = \frac{df}{dK}\sqrt{|\det(\gamma_{\mu\nu})|}. \tag{5.30}$$

We thus find (see Ref. 10):

$$L_{\widetilde{H}} = \pi K(j^2\gamma) - \widetilde{H}(\pi,\gamma_{\mu\nu}). \tag{5.31}$$

A further step allows us to show the equivalence of $L_{\widetilde{H}}$ with an Einstein–Hilbert Lagrangian describing the interaction of the new metric $g_{\alpha\beta}$ defined by setting

$$g_{\alpha\beta} \equiv [\pi/\sqrt{|\det(\gamma_{\mu\nu})|}]\gamma_{\alpha\beta}, \tag{5.32}$$

with the scalar field $\Psi \equiv \log|\pi\sqrt{|\det(\gamma^{\mu\nu})|}|$. Remarkably enough, the definition (5.31) is exactly equivalent to the E–E prescription, even if the Legendre transform is essentially different in this case, since a Lagrangian depending only on $K(j^2\gamma)$ cannot satisfy the condition (5.23). The reader can find in Ref. 10 the detailed calculation for the cases $f \equiv [K(j^2\gamma)]^p$ $(p \neq 0,1)$ and for the well-known theory of Weyl.[36] It is worthwhile to remark that the Legendre transformation for Weyl's theory assumes exactly the form of a "gauge-fixing" prescription, and the scalar field Ψ is reabsorbed by the gauge transformation, so that no new field appears after the whole procedure (on this subject, see also Ref. 11).

[1] A. Einstein, *Zur affinen Feldtheorie* (Sitzungsber. Preuss. Akad. Wiss., Berlin, 1923), pp. 137–140.
[2] A. S. Eddington, *The Mathematical Theory of Relativity*, 2nd ed. (Cambridge U.P., Cambridge, UK, 1924).
[3] E. Schrödinger, Proc. R. Ir. Acad. Sec. A **49**, 43 (1943); E. Schrödinger, Proc. R. Ir. Acad. Sec. A **49**, 275 (1944); E. Schrödinger, Proc. R. Ir. Acad. Sec. A **51**, 41 (1946).
[4] J. Kijowski, Gen. Relativ. Gravit. **9** 10, 857 (1978).
[5] M. Ferraris and J. Kijowski, Rend. Sem. Mat. Univ. Politec. Torino, **41** 3, 169 (1983); M. Ferraris and J. Kijowski, Gen. Relativ. Gravit. **14** 2, 165 (1982); M. Ferraris and J. Kijowski, Lett. Math. Phys. **5** 2, 127 (1981).
[6] M. Ferraris and J. Kijowski, Gen. Relativ. Gravit. **14** 1, 37 (1982).
[7] M. Ferraris, in *Proceedings of the Journées Rélativistes 1983*, edited by S. Benenti, M. Ferraris, and M. Francaviglia, (Tecnoprint, Bologna, 1984), pp. 125–153.
[8] P. W. Higgs, Nuovo Cimento **11** 6, 816 (1959).
[9] B. Whitt Phys. Lett. B **145** 3, 176 (1984).
[10] G. Magnano, M. Ferraris, and M. Francaviglia, Gen. Relativ. Gravit. **19** 5, 465 (1987); M. Ferraris, M. Francaviglia, and G. Magnano, Class. Quantum Grav. **5**, L95 (1988).
[11] M. Ferraris, in *Atti del 6° Convegno Nazionale di Relativitá Generale e Fisica della Gravitazione*, edited by R. Fabbri and M. Modugno, (Tecnoprint, Bologna, 1987), pp. 127–136.
[12] K. S. Stelle, Phys. Rev. D **16** 4, 953 (1977); K. S. Stelle, Gen. Relativ. Gravit. **9** 4, 353 (1978).
[13] S. Cecotti, Phys. Lett. B **190**, 86 (1987).
[14] K. Kuchar, J. Math. Phys. **17** 5, 777, 792, 801 (1976).
[15] R. Arnowitt, S. Deser, and C. W. Misner, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), pp. 227–265.
[16] J. Isenberg and J. Nester, in *General Relativity and Gravitation, One Hundred Years after the Birth of Albert Einstein*, edited by A. Held (Plenum Press, New York, 1980), Vol 1, pp. 23–97; J. Sniaticki, Research Paper No. 600, Department of Mathematics and Statistics, University of Calgary, Calgary, 1985.
[17] T. Regge and C. Teitelboim, Ann. Phys. (N.Y.) **88**, 286 (1974).
[18] W. Szczyrba, J. Math. Phys. **22**, 9, 1926 (1984).
[19] M. Ferraris and M. Francaviglia, Atti Sem. Mat. Univ. Modena **37**, 131 (1989)
[20] W. M. Tulczyjew, in *Symposia Mathematica*, (Academic, London, 1974), Vol. XIV, pp. 247–258.
[21] H. Goldschmidt and S. Sternberg, Ann. Inst. Fourier (Grenoble) **23**, 203 (1973).
[22] P. L. Garcia, in *Symposia Mathematica*, (Academic, London, 1974), Vol. XIV, pp. 219–246.
[23] D. Krupka, Folia Fac. Sci. Nat. UJEP Brunensis (Physica) **14**, 1 (1973).
[24] M. Ferraris and M. Francaviglia, in *Differential Geometry and its Applications*, Proceedings, Brno, 24–30 August 1986, edited by D. Krupka and A. Svec; (Brno, Czechoslovakia, and Reidel, Dordrecht, 1987), pp. 31–52.
[25] M. Francaviglia and D. Krupka, Ann. Inst. H. Poincaré **37**, 3, 295 (1982).
[26] M. Ferraris and M. Francaviglia, in *Proceedings of the International Meeting on Geometry and Physics*, edited by M. Modugno (Pitagora Editrice, Bologna, 1983), pp. 43–70; M. Ferraris and M. Francaviglia, in *Proceedings of the Conference on Differential Geometry and its Applications*, Part 2, Geometrical Methods in Physics, edited by D. Krupka (Brno, Czechoslovakia, 1984), pp. 93–117.
[27] P. L. García and J. Muñoz, C. R. Acad. Sci. Paris **301** (I), 639 (1985).
[28] M. Ferraris, "Dual Lagrangian Theories" (in preparation).
[29] J. F. Pommaret, *Differential Galois Theory* (Gordon & Breach, New York, 1983).
[30] D. Krupka, in *Proceedings of the IUTAM–ISIMM Symposium on Modern*

Developments in Analytical Mechanics, Torino, 7–11 June 1982, edited by S. Benenti, M. Francaviglia, and A. Lichnerowicz (Tecnoprint, Bologna, 1983), pp. 197–238.
[31] I. Kolář, J. Geom. Phys. **1** 2, 127 (1984).
[32] P. L. García and J. Muñoz, in *Proceedings of the IUTAM–ISIMM Symposium on Modern Developments in Analytical Mechanics*, edited by S. Benenti, M. Francaviglia, and A. Lichnerowicz (Tecnoprint, Bologna, 1983), pp. 127–147.
[33] M. Ferraris, in *Proceedings of the Conference on Differential Geometry and its Applications*, Part 2, Geometrical Methods in Physics, edited by D. Krupka (Brno, Czechoslovakia, 1984), pp. 61–91.
[34] R. Ragionieri and R. Ricci, Boll. Un. Mat. Ital. B **18**, 119 (1981).
[35] M. F. Atiyah, *Geometry of Yang–Mills Fields*, Lezioni Fermiane 1979 (Academia Nazionale dei Lincei, Pisa, 1979).
[36] H. Weyl, *Raum-Zeit-Materie*, 1st, 2nd, and 3rd eds., (Springer, Berlin, 1918, 1919, 1921); 4th edition (Dover, New York, 1950).

On the nth-order structure of solitonic wormholes

F. Degen
Department of Physics, University of Wisconsin—Milwaukee, Milwaukee, Wisconsin 53201

(Received 30 May 1989; accepted for publication 16 August 1989)

The structure of the equations describing a solitonic wormhole in the Einstein–Yang–Mills–Higgs system for an arbitrary compact and connected gauge group $\mathscr{G}$ and representation of an arbitrary Higgs field Q is studied. The general structure of these equations and its use in deriving the first-order equations, which result by applying the operator $O_N = \lim_{N\to 0} \partial/\partial N$ to the original equations, where N is the lapse function, is discussed. It is also possible to write down the nth-order equations, which are obtained by applying n times the operator O_N, for a certain class of solutions. These nth-order equations are then specialized to the model with $\mathscr{G} = \mathrm{SU}(2)$ and Q in the adjoint representation. These nth-order equations on the background of a non-Abelian zero-order solution, which can be gauge transformed to satisfy the 't Hooft ansatz at the internal infinity of the hole, are solved. The technique used to solve the system is that of harmonic expansion, yielding an algebraic system of equations which can be interpreted as an eigenvalue problem. The eigenvalues are functions of the parameters of the model and zero-order quantities. Thus, depending on the values of these parameters, nontrivial nth-order solutions may exist or not. Those cases in which there exist nontrivial solutions are stated. Only for those cases can it be expected that global non-Abelian solitonic wormholes are found.

I. INTRODUCTION

Several authors[1–3] proposed long ago to consider black holes as solitons. A careful analysis revealed that by making certain mild conjectures, the only classical solution of the Einstein–Maxwell system satisfying all of the conditions that should be required from a soliton candidate in a theory containing gravity is the nonrotating, magnetically charged, extreme Reissner–Nordström wormhole.[1] Extreme wormholes are characterized by the fact that their Hawking temperature is zero and therefore they are semiclassically stable.

Unlike usual solitons, however, a wormhole is a hard object: Its interaction with high-energy gravitons does not tend to zero as the energy of the gravitons grows beyond all bounds. This suggests that soliton creation and annihilation processes will play an important role in graviton scattering for energies larger than the soliton mass. They might therefore change the high-energy behavior of the theory (see, e.g., Refs. 4 and 5).

It is known that pure Yang–Mills (YM) theories do not possess solitons; one has to couple them to Higgs fields (leading to solutions of the 't Hooft–Polyakov type) or gravity. In Ref. 6 Nieuwenhuizen *et al.* found a regular solution of a 't Hooft–Polyakov magnetic monopole in curved space whose metric is asymptotically Reissner–Nordström with a magnetic charge. However, this solution is not of the form of which we are interested: It is not a hole, but a gravitating monopole. Asymptotically, though, it has the desired behavior. In the Einstein–Maxwell case the solitonic solution is the embedded Reissner–Nordström solution mentioned above. The quantum theory of this soliton was studied in several papers by Hajicek.[4,7] Of course, this solution can always be embedded in a theory with a larger gauge group. The main idea is to work only with the domain of outer communication D with respect to all wormholes present. The space-time is assumed to be asymptotically Minkowskian: D covers the whole of the asymptotic region including both timelike infinities; it is bounded by the future and past horizons, which intersect each other at H, the boundary of the holes; and it is totally hyperbolic. Every Cauchy hypersurface for D contains H. The fields at H must therefore be fixed.

The question arises as to whether more extreme solitonic solutions exist if one couples gravity to non-Abelian YM and scalar fields. This question was studied in a series of papers.[8,9] First it was pointed out that the boundary values at internal infinity H of every solitonic solution of the Einstein–Yang–Mills–Higgs (EYMH) system must satisfy a set of equations, called the zeroth-order equations. This zeroth-order set of equations is obtained from the full EYMH equations at the limit as one approaches the internal infinity (for more details see Ref. 9 or Sec. II of the present paper). It turns out that much can be said about the full (global) solutions by considering the zeroth-order system, e.g., about the rigidity of the geometry of the internal infinity of an extreme hole. Especially, the boundary values at H of fields belonging to global solutions must be elements of the set of solutions of the zeroth-order equations.

However, the zeroth-order solutions are not sufficient if one wants to know more details about the global solutions. It might be that different global solutions have the same boundary values at the internal infinity H and thus the same zeroth-order behavior, i.e., they will be indistinguishable in the zeroth-order system. One is therefore led to study the so-called "nth-order equations," which are nothing but the limit of the nth-order derivative of the field equations with respect to the lapse function N as one approaches the internal infinity H (N measures the radial distance in a Cauchy surface from the compact, two-dimensional surface H defined by $N=0$).

The set of solutions of the nth-order equations contains the allowed boundary values of the nth-order derivatives of the fields at $N=0$, building up a soliton. The first-order

 0022-2488/90/020388-13$03.00

system was studied for a specific model and it was found that nontrivial (i.e., nonzero) solutions of this system existed (the nth-order system is a linear system; hence the zero solution is always a trivial solution). Some of the solutions existed only if the parameters of the model satisfied certain relations. The existence of these nontrivial first-order solutions might indicate the existence of different solitonic solutons having the same zeroth order, i.e., the same boundary values at the internal infinity H.

It is clear that some of the first-order solutions correspond to the radial change of a given global solution having the zeroth-order value used as "background," but some other first-order solutions might indicate that different families of solutions are branching off from a common zeroth-order value. Here we are faced with a problem analogous to that of linearization stability in the case of perturbations around some global solution: The existence of nontrivial first-order solutions (apart from those describing the "radial" behavior) does not guarantee that there is in fact a new family of solutions branching off. As the equations split into evolution equations along N and evolutive constraints on $N=$ const surfaces, which are propagated by the evolution equations,[10] an analysis analogous to that given in Ref. 11 might be performed, giving clues as to which first-order solutions are tangent to a curve of global solutions. A recent article by Bartnick and McKinnon[12] reports numerical calculations which indicate the existence of a smooth nonsingular solution in an EYM model, which could be tangent to the first-order solution described in Ref. 8: Their particlelike solution is regular everywhere and asymptotically Schwarzschild. Only near internal infinity could it be approximated by one of our solutions.

In this paper we will derive the general first- and nth-order equations for special cases. This is easily done if one studies what we shall call the "N structure" of the full field equations, i.e., the way in which the lapse function N enters the field equations and thus determines which terms will not vanish in the limit $N\to 0$. This allows us to prove, among other things, e.g., that the $N=0$ surface corresponds to a Killing horizon.[13]

Specializing the first-order equations to a definite zeroth-order solution, i.e., inserting for the zeroth-order quantities (such as the boundary values of the electric field E, the magnetic field B, the metric g_{AB}, etc.) appearing in the first-order equations one of the solutions found in Refs. 9 and 14, one obtains a definite set of equations which can be solved by expanding the fields in appropriate harmonics. In Ref. 8 this was done for a SU(2) model using as the zeroth-order solution the one realized by the embedded Reissner–Nordström solution.

One can do the same for the zeroth-order solution having a non-Abelian monopole structure. If one uses the nth-order equations, the conditions for the existence of nontrivial nth-order solutions will lead to equations relating the order-counting parameter n and the parameters of the model.

It is clear that the nth-order equations can also be obtained in another way: Expand all fields in terms of the parameter N, insert the expansions into the field equations, and consider separately every order of N.

In Ref. 8 it was proposed to construct extreme solutions of the EYMH system in a series in the parameter N. The different coefficients are nothing but the solutions of our nth-order equations. We are, of course, particularly interested in source-free global solutions, i.e., without any charge or matter distributions between static internal infinity and i^0. The problem consists in finding which contribution belongs to which class of solutions.

This article is organized as follows. In Sec. II we fix the notation and restate the characteristics that we require from a soliton candidate. In Sec. III we write explicitly the EYMH system in such a way that the relevant N structure is evident. This helps us in Sec. IV to derive the first-order equations and a special case of the nth-order equations. In Sec. V we specialize our equations to the case with the gauge group $\mathscr{G}=\mathrm{SU}(2)$ and insert the non-Abelian zeroth order solution found in Ref. 14, which together with the Abelian solutions derived in Ref. 9 and analyzed in Ref. 8 seem to be all the zeroth-order solutions. This is in all respects analogous to a perturbative calculation on a given background. In Secs. VI and VII we expand the fields in the appropriate way and insert this expansion into the equations. This leads to a linear system of equations with the characteristics of an eigenvalue problem. The eigenvalue is given in terms of zeroth-order quantities and the parameters of the model. We then discuss in Sec. VIII the existence of nontrivial solutions to this system for each of the three possible cases discussed in Ref. 14.

II. THE MODEL

In this section we want to fix our notation and give a brief summary of the model. More details can be found in Ref. 9. Our Lagrangian is

$$L=L_G+L_M,$$

where L_G and L_M are given by

$$L_G=\frac{1}{16\pi\gamma^2}\int d^3x\sqrt{-\hat{g}}\,(\hat{R}-2\Lambda)$$

and

$$L_M=\frac{1}{4\pi}\int d^3x\sqrt{-\hat{g}}\times\left[\tfrac{1}{4}(G_{\mu\nu},G^{\mu\nu})_g+\tfrac{1}{2}(D_\mu Q,D^\mu Q)_q+V(Q)\right].$$

For later convenience we denote four-dimensional quantities with a caret. Elsewhere we use the same notation as in Ref. 9: $\hat{g}_{\mu\nu}$ is the metric of space-time, $\hat{g}\doteq\det(\hat{g}_{\mu\nu})$, γ^2 is Newton's constant, and Λ is the cosmological constant. The gauge field strengths $G_{\mu\nu}$ are given by

$$G_{\mu\nu}=\partial_\mu W_\nu-\partial_\nu W_\mu-[W_\mu,W_\nu],$$

where the Lie-algebra-valued potential W_μ transforms under the gauge transformations $U(x)\in\mathscr{G}$, with $\mathscr{G}$ being the gauge group, as

$$W'_\mu=\mathrm{ad}(U)W_\mu-U^{-1}\partial_\mu U.$$

The gauge-covariant derivative of a Lie-algebra-valued field

$A(x)$ that transforms homogeneously under gauge transformations U is

$$D_\mu A = \partial_\mu A - \mathrm{ad}(W_\mu)A.$$

The gauge transformation $U(x)$ changes the scalar field Q, which is assumed to be an element of a linear space R_q in which a representation $\mathscr{D}$ of the group $\mathscr{G}$ is realized, as follows:

$$Q'(x) = \mathscr{D}(U(x))Q(x).$$

The corresponding gauge-covariant derivative is

$$D_\mu Q = \partial_\mu Q - \mathscr{D}(W_\mu)Q.$$

We use the following conventions: The signature of the metric $\hat{g}_{\mu\nu}$ is $+2$, the curvature tensor $\hat{R}^{\mu}{}_{\nu\rho\sigma}$ is defined by

$$\hat{R}^{\mu}{}_{\nu\rho\sigma} = \partial_\rho \hat{\Gamma}^{\mu}_{\nu\sigma} - \partial_\sigma \hat{\Gamma}^{\mu}_{\nu\rho} + \hat{\Gamma}^{\mu}_{\kappa\rho}\hat{\Gamma}^{\kappa}_{\nu\sigma} - \hat{\Gamma}^{\mu}_{\kappa\sigma}\hat{\Gamma}^{\kappa}_{\nu\rho},$$

the Ricci tensor $\hat{R}_{\mu\nu}$ is defined by

$$\hat{R}_{\mu\nu} = \hat{R}^{\rho}{}_{\mu\rho\nu},$$

and the Ricci scalar $\hat{R}$ is defined by

$$\hat{R} = \hat{g}_{\mu\nu}\hat{R}_{\mu\nu}.$$

The gauge group is assumed to be a compact and connected Lie group whose Lie algebra g admits a real, bilinear, symmetric form $(,)_g$ which is positive definite, constant on space-time, and invariant under the action of the group.

We also assume that a real, symmetric, bilinear form exists on R_q, which we denote by $(,)_q$: It is positive definite, constant on space-time, and invariant under the action of the group.

The representation $\mathscr{D}$ and the two forms $(,)_q$ and $(,)_g$ determine uniquely a g-valued bilinear, skew-symmetric form $\omega(,)$ according to

$$(Q_1,\mathscr{D}(A)Q_2)_q = (\omega(Q_1,Q_2),A)_g.$$

The function $V(Q)$ is a real function on R_q satisfying the following conditions:

$$\partial_\mu V = \left(\frac{\partial V}{\partial Q},\partial_\mu Q\right)_q$$

and

$$\left(\frac{\partial V}{\partial Q},\mathscr{D}(A)Q\right)_q = 0, \quad \forall A\in g, \quad Q\in R_q,$$

which guarantee constancy on space-time and invariance under the action of the group. Using the properties of the linear forms, the variation of the Lagrangian L leads to the following field equations:

$$\hat{R}_{\mu\nu} - \tfrac{1}{2}\hat{R}\hat{g}_{\mu\nu} = 8\pi\gamma^2\overline{T}_{\mu\nu}, \tag{2.1}$$

$$(1/\sqrt{-\hat{g}})D_\nu(\sqrt{-\hat{g}}G^{\mu\nu}) - \omega(D^\mu Q,Q) = 0, \tag{2.2}$$

$$\frac{1}{\sqrt{-\hat{g}}}D_\mu(\sqrt{-\hat{g}}D^\mu Q) - \frac{\partial V}{\partial Q} = 0. \tag{2.3}$$

Equations (2.1)–(2.3) are covariant with respect to general coordinate and gauge transformations. The effective energy-momentum tensor $\overline{T}_{\mu\nu}$ is given by

$$\overline{T}_{\mu\nu} = T_{\mu\nu} - (\Lambda/8\pi\gamma^2)\hat{g}_{\mu\nu}$$

and

$$\begin{aligned} T^{\mu\nu} &= -(2/\sqrt{-\hat{g}})(\delta L_M/\delta\hat{g}_{\mu\nu}) \\ &= (1/4\pi)\big[(G^{\mu\rho},G^{\nu}{}_{\rho})_g \\ &\quad - \tfrac{1}{4}\hat{g}^{\mu\nu}(G_{\rho\sigma},G^{\rho\sigma})_g + (D_\mu Q,D^\mu Q)_q \\ &\quad - \tfrac{1}{2}\hat{g}^{\mu\nu}(D_\rho Q,D^\rho Q)_q - \hat{g}^{\mu\nu}V(Q)\big]. \end{aligned} \tag{2.4}$$

Our aim is to find out whether the EYMH system of equations (2.1)–(2.3) has more than one global solitonic solution, i.e., the embedded Reissner–Nordström solution. These solutions must satisfy a set of conditions if they are to be considered as solitons. These conditions are listed in full detail in Ref. 9 and here we give a brief summary:

(i) The solution $(\mathscr{M},\hat{g}_{\mu\nu},W_\mu,Q)$ is static. In fact, this condition can be weakened since we only require that the solution is asymptotically static as one approaches the internal boundary.

(ii) Here $(\mathscr{M},\hat{g}_{\mu\nu},W_\mu,Q)$ is maximal.

(iii) If S is a hypersurface in $(\mathscr{M},\hat{g}_{\mu\nu})$ that is orthogonal to the timelike Killing vector and inextendable, then S is a Cauchy surface for $(\mathscr{M},\hat{g}_{\mu\nu})$ and further, it is complete with respect to the distance function associated with the positive definite metric induced by $\hat{g}_{\mu\nu}$ on S.

(iv) The space-time $(\mathscr{M},\hat{g}_{\mu\nu})$ is asymptotically flat and the ADM energy is finite.

(v) From conditions (i) and (iii), together with a further technical assumption regarding the function g_{00}, it follows that the space-time metric can be written in the form

$$ds^2 = -N^2\,dt^2 + \rho^2\,dN^2 + g_{AB}\,dx^A\,dx^B, \tag{2.5}$$

where $A,B = 2,3, N>0$ in $\mathscr{H}$ and where $\mathscr{H}$ is the nonempty set of regions in S in which g_{00} has no critical points and the surfaces $g_{00} = \text{const}$ are compact. Further, $\rho = \rho(N,x^2,x^3) > 0$, $g_{AB} = g_{AB}(N,x^2,x^3)$, and resulting from condition (iii),

$$\lim_{N\to 0}\rho^{-1} = 0.$$

We will denote the covariant derivative with respect to the four-dimensional metric $\hat{g}_{\mu\nu}$ with $\hat{\nabla}$ and the covariant derivative with respect to the two-dimensional metric g_{AB} with ∇.

(vi) The quantities $\rho, g_{AB}, W_0, W_A, Q$ are smooth (C^∞) functions of the coordinates N, x^2, x^3 in $\mathscr{H}$. The derivatives of g_{AB}, W_A, Q of order $0,1,\ldots,n+2$ and of $(N\rho)^{-1}\partial_1 W_0$ of order $0,1,\ldots,n+1$ with respect to N are continuous functions of N and smooth functions of x^2, x^3 at $N=0$. For the pure Einstein or Einstein–Maxwell system, the smoothness follows from much weaker conditions, e.g., some curvature scalars must be bounded.

We have restricted the gauge of W_μ and Q as follows:

$$\partial_0 W_\mu = 0 = \partial_0 Q \tag{2.6}$$

and

$$W_1 = 0. \tag{2.7}$$

The completeness of S guarantees that the hole is of the extreme type. In Ref. 1 it was shown that in the Einstein–Maxwell system the only possible soliton candidate satisfying both classical and semiclassical conditions were the extreme holes. We point out that we do not assume the existence of a regular horizon. However, it was shown that solutions satisfying conditions (i)–(vi) always have a Killing horizon.[13]

In order for W_μ to have finite components in an orthonormal tetrad, the following must be true:

$$\lim_{N\to 0} W_0 = 0. \tag{2.8}$$

The radial (equipotential surface orthogonal) electric field is

$$E = (1/N\rho) G_{10} \tag{2.9}$$

and the radial magnetic field is

$$B = \tfrac{1}{2} e^{AB} G_{AB}, \tag{2.10}$$

where $e^{AB} \doteq \epsilon^{AB}/\sqrt{g}$, ϵ^{AB} is the two-dimensional Levi-Civita symbol ($\epsilon^{23} = 1$), and $g \doteq \det(g_{AB})$.

III. THE GLOBAL EQUATIONS

In this section we write explicitly the EYMH system (2.1)–(2.3) in such a way that all quantities appearing in it have a finite limit for $N\to 0$. This will allow us to recognize the N structure of the system, i.e., the way in which the system depends on the lapse function N. We write the equations in a form that is covariant in the coordinates x^2, x^3. The projection method needed for this has been worked out by Israel.[10] We define a new quantity w:

$$w \doteq W_0/N, \quad w \in g, \tag{3.1}$$

which has a regular limit for $N\to 0$. The radial electric field E is related to w through

$$E = (N/N\rho)\partial_1 w + (1/N\rho) w. \tag{3.2}$$

We also use the following relation:

$$\partial_1 \sqrt{g} = \tfrac{1}{2}\sqrt{g}\, g^{AB} \partial_1 g_{AB}. \tag{3.3}$$

With relations (3.2) and (3.3), together with the metric (2.5), the gauge conditions (2.6) and (2.7), and Eq. (2.8) we can write the equations in a form that will clearly reveal their N structure, i.e., the special way in which the lapse function enters into the equations.

We begin with the Einstein equations (2.1) in the projected form.[10,9] The first equation is obtained by basically taking the traceless part of the A, B components:

$$\begin{aligned}
&\frac{N^2}{(N\rho)^2} g^{AC}\partial_1{}^2 g_{CB} - \frac{N^2}{(N\rho)^2} g^{AK} g^{CL} (\partial_1 g_{KL})(\partial_1 g_{CB}) + \frac{2N}{(N\rho)^2} g^{AC}\partial_1 g_{CB} \\
&\quad + \frac{1}{2}\frac{N^2}{(N\rho)^2} g^{KL}(\partial_1 g_{KL}) g^{AC}(\partial_1 g_{CB}) - \frac{2N^2}{(N\rho)^3}(\partial_1 N\rho) g^{AC}\partial_1 g_{CB} - R\delta^A{}_B + \frac{2}{N\rho}(N\rho);^A{}_B \\
&\quad = 8\pi\gamma^2(\overline{T}^\mu{}_\mu \delta^A{}_B - 2\overline{T}^A{}_B) \\
&\quad = 4\gamma^2\Big[g^{AK}\delta^L{}_B - \frac{1}{2}\delta^A{}_B g^{AB}\Big]\Big[(D_K w, D_L w)_g - \frac{N^2}{(N\rho)^2}(\partial_1 W_K, \partial_1 W_L)_g + (D_K Q, D_L Q)_q\Big] \\
&\quad - 2\gamma^2\delta^A{}_B\Big[\frac{1}{(N\rho)^2}(N\partial_1 w, N\partial_1 w)_g + \frac{2}{(N\rho)^2}(N\partial_1 w, w)_g \\
&\quad + \frac{1}{(N\rho)^2}(w,w)_g + (B,B)_g - g^{AB}(D_A Q, D_B Q)_q + 2V(Q) + \frac{\Lambda}{\gamma^2}\Big].
\end{aligned} \tag{3.4}$$

The following equation is a combination of the (0,0) component and the trace:

$$\begin{aligned}
&\frac{2}{(N\rho)^2} + \frac{N}{(N\rho)^2} g^{AB}\partial_1 g_{AB} - \frac{2N}{(N\rho)^3}(\partial_1 N\rho) \\
&\quad = 8\pi\gamma^2(\overline{T}^\mu{}_\mu - 2\overline{T}^0{}_0) \\
&\quad = 4\gamma^2\Big[(\mathscr{D}(w)Q, \mathscr{D}(w)Q)_q + \frac{1}{2}\frac{N^2}{(N\rho)^2} g^{AB}(\partial_1 W_A, \partial_1 W_B)_g + \frac{1}{2} g^{AB}(D_A w, D_B w)_g + \frac{1}{2}\frac{1}{(N\rho)^2}(N\partial_1 w, N\partial_1 w)_g \\
&\quad + \frac{1}{(N\rho)^2}(N\partial_1 w, w)_g + \frac{1}{2}\frac{1}{(N\rho)^2}(w,w)_g + \frac{1}{2}(B,B)_g - \frac{1}{2}\Big(2V(Q) + \frac{\Lambda}{\gamma^2}\Big)\Big].
\end{aligned} \tag{3.5}$$

As noted by Israel,[10] Eqs. (3.4) and (3.5) form a complete system for determining the evolution of g_{AB} and $N\rho$ as functions of N. The following equations are constraints which, once they are satisfied on one $N =$ const hypersurface, are satisfied identically. The tangential stresses T^{AB} can be assigned arbitrarily over the three-space; the conservation identities $\hat{\nabla}_\mu T^{\mu\nu} = 0$ are satisfied automatically if the normal stresses T^{11} and T^{A1} are determined by the following two equations:

$$\frac{N^2}{4(N\rho)^2}[g^{AK}g^{BL}(\partial_1 g_{AB})(\partial_1 g_{KL}) - g^{AB}(\partial_1 g_{AB})g^{KL}(\partial_1 g_{KL})] - \frac{N}{(N\rho)^2} g^{AB}\partial_1 g_{AB} + R$$
$$= -16\pi\gamma^2 \overline{T}^1_1 = -4\gamma^2\Big[-\frac{1}{2}g^{AB}(D_A Q, D_B Q)_q - \frac{1}{2}\frac{1}{(N\rho)^2}(N\partial_1 w, N\partial_1 w)_g - \frac{1}{(N\rho)^2}(N\partial_1 w, w)_g$$
$$-\frac{1}{2}\frac{1}{(N\rho)^2}(w,w)_g - \frac{1}{2}(B,B)_g + \frac{1}{2}(\mathscr{D}(w)Q, \mathscr{D}(w)Q)_q + \frac{1}{2}\frac{N^2}{(N\rho)^2}(\partial_1 Q, \partial_1 Q)_q$$
$$+\frac{1}{2}g^{AB}(D_A w, D_B w)_g + \frac{1}{2}\frac{N^2}{(N\rho)^2}g^{AB}(\partial_1 W_A, \partial_1 W_B)_g - \frac{1}{2}\Big(2V(Q) + \frac{\Lambda}{\gamma^2}\Big)\Big], \tag{3.6}$$

$$\partial_A(Ng^{KL}\partial_1 g_{KL}) - \nabla_B(Ng^{BL}\partial_1 g_{AL}) - (N/N\rho)g^{KL}(\partial_1 g_{KL})(\partial_A N\rho) + (N/N\rho)g^{BL}(\partial_1 g_{AL})(\partial_B N\rho) - (2/N\rho)\partial_A N\rho$$
$$= -16\pi\gamma^2 N\overline{T}_{1A} = -4\gamma^2[-(N\partial_1 w, D_A w)_g - (w, D_A w)_g + Ng^{KL}(\partial_1 W_L, e_{AK}B)_g + N(\partial_1 Q, D_A Q)_g]. \tag{3.7}$$

The electric equation [Eq. (2.2) with $\mu = 0$] has the following form:

$$-\frac{N^2}{(N\rho)^2}\partial_1{}^2 w - \frac{N}{(N\rho)^2}g^{AB}\partial_1 g_{AB}\Big(\frac{1}{2}N\partial_1 w + \frac{1}{2}w\Big) + \frac{N^2}{(N\rho)^3}(\partial_1 N\rho)\partial_1 w - \frac{2N}{(N\rho)^2}\partial_1 w$$
$$-\frac{N}{(N\rho)^3}(\partial_1 N\rho)w - \frac{1}{N\rho}g^{AB}(\partial_A N\rho)D_B w - \frac{1}{\sqrt{g}}D_A(\sqrt{g}g^{AB}D_B w) - \omega(Q, \mathscr{D}(w)Q) = 0. \tag{3.8}$$

For the magnetic equation [Eq. (2.2) with $\mu = 2,3$] we obtain, after carrying out the derivative of the first two terms,

$$-\frac{1}{2}\frac{N^2}{(N\rho)^2}g^{KL}(\partial_1 g_{KL})g^{AB}\partial_1 W_B - \frac{N^2}{(N\rho)^2}g^{AB}\partial_1{}^2 W_B - \frac{2N}{(N\rho)^2}g^{AB}\partial_1 W_B + \frac{N^2}{(N\rho)^3}(\partial_1 N\rho)g^{AB}(\partial_1 W_B)$$
$$+\frac{N^2}{(N\rho)^2}g^{AK}g^{BL}(\partial_1 g_{KL})(\partial_1 W_B) + \frac{e^{AB}}{N\rho}(\partial_B N\rho)B + e^{AB}D_B B + g^{AB}\mathrm{Ad}(w)D_B w + g^{AB}\omega(Q, D_B Q) = 0. \tag{3.9}$$

The scalar equation (2.3)

$$\frac{N^2}{(N\rho)^2}\partial_1^2 Q + \frac{2N}{(N\rho)^2}\partial_1 Q + \frac{1}{2}\frac{N^2}{(N\rho)^2}g^{AB}(\partial_1 g_{AB})\partial_1 Q - \frac{N^2}{(N\rho)^2}(\partial_1 N\rho)\partial_1 Q$$
$$+\frac{1}{N\rho}(\partial_A N\rho)g^{AB}D_B Q + \frac{1}{\sqrt{g}}D_A(\sqrt{g}g^{AB}D_B Q) - \frac{\partial V}{\partial Q} - \mathscr{D}(w)\mathscr{D}(w)Q = 0. \tag{3.10}$$

Equation (2.2) with $\mu = 1$ is a dependent equation: It takes the form

$$-\frac{N}{(N\rho)^3}(\partial_A N\rho)g^{AB}\partial_1 W_B$$
$$+\frac{N}{(N\rho)^2}\frac{1}{\sqrt{g}}D_A(\sqrt{g}g^{AB}\partial_1 W_B)$$
$$+\frac{1}{(N\rho)^2}\mathrm{Ad}(w)(N\partial_1 w + w)$$
$$+\frac{N}{(N\rho)^2}\omega(Q, \partial_1 Q) = 0. \tag{3.11}$$

The magnetic field B is given by the gauge potentials W_A through Eq. (2.10). Hence, after inspection of Eqs. (3.4)–(3.10) we can write the equations schematically as

$$\mathscr{S}^A_c(\Psi)N^2\partial_1^2\Psi^c + \mathscr{F}^A_{cg}(\Psi)(N\partial_1\Psi^c)(N\partial_1\Psi^g)$$
$$+\mathscr{L}^A_c(\Psi)(N\partial_1\Psi^c) + \mathscr{M}^A(\Psi) = 0, \tag{3.12}$$

where Ψ^c is a shorthand notation for the fields $(N\rho, g_{AB}, w, W_A, Q)$ and we sum over repeated indices.

Here $\mathscr{S}^A_c$, $\mathscr{F}^A_{cg}$, $\mathscr{L}^A_c$, $\mathscr{M}^A$ represent differential operators depending on Ψ, but not containing any factors of N nor derivatives with respect to N.

IV. THE *n*th-ORDER EQUATIONS

In this section we study the behavior of Eq. (3.12) and its derivatives with respect to N in the limit $N\to 0$. As noted before, all quantities appearing in Eq. (3.12) have regular limit for $N\to 0$; all the explicit dependence on N, as well as the derivatives with respect to N, are explicitly shown. Of course, the operators $\mathscr{S}^A_c$, $\mathscr{F}^A_{cg}$, $\mathscr{L}^A_c$, $\mathscr{M}^A$ implicitly depend on N through Ψ, which is a shorthand notation for all the fields.

Therefore, it is not difficult to derive the limit of Eq. (3.12) and its derivatives with respect to N in the limit $N\to 0$. Thus, e.g.,

$$\lim_{N\to 0} \mathscr{M}^A(\Psi) = \mathscr{M}^A(\overset{(0)}{\Psi}) = 0, \tag{4.1}$$

where we have defined

$$\overset{(0)}{\Psi} \doteq \lim_{N\to 0} \Psi, \tag{4.2}$$

which are the zeroth-order equations, i.e., the limit of Eq. (3.12) for $N\to 0$: Their solutions are the possible boundary values of the global extreme solutions of the EYMH system (2.1)–(2.3). This zeroth-order system was studied in Ref. 9.

Similarly, one can derive the first-order equations, i.e., the limit of the first derivative with respect to N of Eq. (3.12). These equations must be satisfied by the normal derivatives of extreme hole solutions of the EYMH system and are given by

$$\lim_{N\to 0} [\mathscr{L}^A_c(\Psi) + \mathscr{M}^A_{,c}]\overset{(1)}{\Psi^c} = \left[\mathscr{L}^A_c(\overset{(0)}{\Psi}) + \mathscr{M}^A_{,c}(\overset{(0)}{\Psi})\right]\overset{(1)}{\Psi^c} = 0,$$

where we have defined

$$\overset{(1)}{\Psi} \doteq \lim_{N\to 0} \partial_1 \Psi \tag{4.3}$$

and

$$\mathscr{M}^A_{,c} \doteq (\delta/\delta\Psi^c)\mathscr{M}^A. \tag{4.4}$$

One can even write the nth-order equations for the special case in which the first-, second-,...,$(n-1)$-order solutions vanish (but the zeroth-order solution does not). For the case $n=1$ these equations clearly reduce to the general first-order equations. Using the Leibniz rule and the N structure we obtain

$$\lim_{N\to 0}[n(n-1)\mathscr{S}^A_c(\Psi) + n\mathscr{L}^A_c(\Psi) + \mathscr{M}^A_{,c}(\Psi)]\overset{(n)}{\Psi^c}$$
$$= \left[n(n-1)\mathscr{S}^A_c(\overset{(0)}{\Psi}) + \mathscr{L}^A_c(\overset{(0)}{\Psi}) + \mathscr{M}^A_{,c}(\overset{(0)}{\Psi})\right]\overset{(n)}{\Psi^c} = 0, \tag{4.5}$$

where we define $\overset{(n)}{\Psi^c}$ as follows:

$$\overset{(n)}{\Psi^c} \doteq \lim_{N\to 0} \partial_1^n \Psi^c \tag{4.6}$$

and

$$\overset{(i)}{\Psi^c} = 0, \quad i = 1,2,\ldots,n-1,$$

i.e., we only consider nth-order solutions which have vanishing lower-order $(1,2,\ldots,n-1)$ support.

We are now in a position to write the nth-order equations for the case of a nonvanishing zeroth-order solution, but vanishing first-, second-,...,$(n-1)$-order solutions. As mentioned above, for $n=1$ these are the general first-order equations. We begin with Einstein's equations. From now on we will replace the superscript (n) by an overtilde. The order that is meant should always be clear from the context. In addition, we will use the following abbrevations:

$$\lim_{N\to 0} N\rho \doteq G, \tag{4.7}$$

$$\lim_{N\to 0} \partial_1{}^n N\rho \doteq \widetilde{G}, \tag{4.8}$$

$$\lim_{N\to 0} g^{AB}\,\partial_1 g_{AB} \doteq \kappa, \tag{4.9}$$

$$\lim_{N\to 0} g^{AB}\,\partial_1{}^n g_{AB} \doteq \tilde{\kappa}. \tag{4.10}$$

Equation (3.4) leads to

$$[n(n+1)/G^2]\, g^{AC}\tilde{g}_{CB} - \widetilde{R}\delta^A{}_B + (2/G)\widetilde{G}_;{}^A{}_B$$
$$= 8\pi\gamma^2(\widetilde{\widetilde{T}}{}^\mu{}_\mu\delta^A{}_B - 2\widetilde{\widetilde{T}}{}^A{}_B). \tag{4.11}$$

Here $\widetilde{R}$ is related to $\tilde{\kappa}$ via

$$\widetilde{R} = \Delta\tilde{\kappa} + g^{AK}g^{BL}\nabla_B\nabla_A\tilde{g}_{KL} - \tfrac{1}{2}\tilde{\kappa}R. \tag{4.12}$$

For Eq. (3.5) we obtain the nth order,

$$-[2(n+2)/G^3]\widetilde{G} + (n/G^2)\tilde{\kappa} = 8\pi\gamma^2(\widetilde{\widetilde{T}}{}^\mu{}_\mu - 2\widetilde{\widetilde{T}}{}^0{}_0), \tag{4.13}$$

for Eq. (3.6) we obtain

$$-(n/G^2)\tilde{\kappa} + \widetilde{R} = -16\pi\gamma^2\widetilde{\widetilde{T}}{}^1{}_1, \tag{4.14}$$

and for Eq. (3.7) we obtain

$$n(g^{KL}\delta^B{}_A - g^{BL}\delta^K{}_A)\nabla_B\tilde{g}_{KL} - (2/G)\nabla_A\widetilde{G}$$
$$= -16\pi\gamma^2 \lim_{N\to 0} N\widetilde{\widetilde{T}}_{1A}. \tag{4.15}$$

Equation (4.12) can easily be derived from the well-known relation

$$\delta R = \nabla_L\delta\Gamma^L{}_{AB} - \nabla_B\delta\Gamma^L{}_{AL},$$

which is, as the rest of these equations, valid as an nth-order equation only for vanishing first, second,...,$(n-1)$ order, respectively, for $n=1$. We have used the zeroth-order result:

$$\partial_A G = 0. \tag{4.16}$$

It is easy to show from the explicit form of T^μ_ν that the following is true:

$$\widetilde{\widetilde{T}}{}^0{}_0 = \widetilde{\widetilde{T}}{}^1{}_1. \tag{4.17}$$

We now define the new variable

$$\tilde{\gamma}_{KL} \doteq (\delta^A{}_K\delta^B{}_L - \tfrac{1}{2}g^{AB}g_{KL})\tilde{g}_{AB}. \tag{4.18}$$

We also note that as a result of the definition of B [(2.10)] the following relation holds:

$$\widetilde{B} = -\tfrac{1}{2}\tilde{\kappa}B + e^{AB}D_A\widetilde{W}_B. \tag{4.19}$$

After some algebra the symmetric traceless part of Eq. (4.11), together with Eqs. (4.13), (4.14), and zeroth-order relations such as

$$\lim_{N\to 0} \mathscr{D}(w)Q = 0, \tag{4.20}$$

$$\lim_{N\to 0} D_A w = 0, \tag{4.21}$$

yields, after some algebra,

$$[n(n+1)/G^2]\tilde{\gamma}_{AB} + (2/G)(\nabla_A\nabla_B\widetilde{G}$$
$$-\tfrac{1}{2}g_{AB}\Delta\widetilde{G}) - 2\gamma^2 g^{KL}\tilde{\gamma}_{AB}(D_KQ,D_LQ)$$
$$= -4\gamma^2(\delta^K{}_A\delta^L{}_B + \delta^K{}_B\delta^L{}_A - g^{KL}g_{AB})$$
$$\cdot[(D_K\widetilde{Q},D_LQ)_q - (\mathscr{D}(\widetilde{W}_K)Q,D_LQ)_q]. \tag{4.22}$$

Inserting Eq. (4.14) into (4.11), using (4.17), and taking the trace yields

$$(2/G)\Delta\tilde{G} = -[n(n-1)/G^2]\tilde{\kappa}, \quad (4.23)$$

where Δ denotes the covariant Laplacian for g_{AB}. Equation (4.24) follows directly from Eq. (4.13), using zeroth-order properties such as (4.16), (4.20), and (4.21):

$$-\frac{2(n+2)}{G^3}\tilde{G} + \frac{n}{G^2}\tilde{\kappa} + 4\gamma^2\left[\frac{1}{G^3}\tilde{G}(w,w)_g + \frac{1}{2}\tilde{\kappa}(B,B)_g\right] = 4\gamma^2\left[\frac{n+1}{G^2}(\tilde{w},w)_g + e^{AB}(D_A\tilde{W}_B,B)_g - \left(\frac{\delta V}{\delta Q},\tilde{Q}\right)_q\right]. \quad (4.24)$$

From Eqs. (4.14) and (4.12) and the zeroth-order relations given in Ref. 9 we obtain

$$-\frac{1}{2}\Delta\tilde{\kappa} + \nabla_B\nabla_A\tilde{\gamma}^{AB} - \frac{n-1}{G^2}\tilde{\kappa} - \frac{2\gamma^2}{G^2}(w,w)_g\left(\tilde{\kappa} - \frac{2}{G}\tilde{G}\right) = 4\gamma^2\Big[(D_A\tilde{Q},D_AQ)_q + e^{AB}(\tilde{W}_B,D_AB)_g + \frac{n+1}{G^2}(\tilde{w},w)_g + e^{AB}(D_A\tilde{W}_B,B)_g + \left(\frac{\partial V}{\partial Q},\tilde{Q}\right)_q\Big], \quad (4.25)$$

where

$$\tilde{\gamma}^{AB} \doteq g^{AC}g^{BL}\tilde{\gamma}_{CL}. \quad (4.26)$$

Equation (4.27) is the nth-order equation of (4.15):

$$(n/2)\tilde{\kappa}_{;A} - n\nabla_B\tilde{\gamma}^B{}_A - (2/G)\nabla_A\tilde{G} = 4\gamma^2[(w,D_A\tilde{w})_g - ne_A{}^K(\tilde{W}_K,B)_g - n(\tilde{Q},D_AQ)_q]. \quad (4.27)$$

We now come to the YMH equations. The electric equation (3.8) has the following nth-order [remember that we assume that the first, second,..., $(n-1)$th order vanish]:

$$\frac{n(n+1)}{G^2}\tilde{w} + \frac{n}{2G^2}w\tilde{\kappa} - \frac{n}{G^3}w\tilde{G} + \mathscr{D}^A\mathscr{D}_A\tilde{w} - g^{AB}\,\mathrm{Ad}(\mathscr{D}_A\tilde{W}_B)w + \omega(Q,\mathscr{D}(\tilde{w})Q) + \omega(Q,\mathscr{D}(w)\tilde{Q}) = 0, \quad (4.28)$$

where $\mathscr{D}_A$ stands for the gauge and coordinate covariant derivative (i.e., it is obtained from the gauge covariant derivative by using ∇_A instead of ∂_A). The nth-order magnetic equation is

$$[n(n+1)/G^2]g^{AB}\tilde{W}_B - (1/G)e^{AB}B\nabla_B\tilde{G} + \tfrac{1}{2}e^{AB}B\nabla_B\tilde{\kappa} - e^{AB}e^{KL}\mathscr{D}_B\mathscr{D}_K\tilde{W}_L + e^{AB}\,\mathrm{Ad}(\tilde{W}_B)B - g^{AB}\,\mathrm{Ad}(w)\mathscr{D}_B\tilde{w} + g^{AB}\,\mathrm{Ad}(w)\,\mathrm{Ad}(\tilde{W}_B)w + (\tilde{\gamma}^{AB} - \tfrac{1}{2}g^{AB}\tilde{\kappa})\omega(Q,\mathscr{D}_BQ) - g^{AB}\omega(\tilde{Q},\mathscr{D}_BQ) - g^{AB}\omega(Q,\mathscr{D}_B\tilde{Q}) + g^{AB}\omega(Q,\mathscr{D}(\tilde{W}_B)Q) = 0. \quad (4.29)$$

The nth-order scalar equation is

$$\frac{n(n+1)}{G^2}\tilde{Q} + \frac{1}{G}g^{AB}\nabla_B\tilde{G}\mathscr{D}_BQ - \widetilde{\frac{\partial V}{\partial Q}} - \mathscr{D}(w)\mathscr{D}(\tilde{w})Q - \mathscr{D}(w)\mathscr{D}(w)\tilde{Q} - \frac{1}{2}\tilde{\kappa}\frac{\partial V}{\partial Q} - 2g^{AB}\mathscr{D}(\tilde{W}_A)\mathscr{D}_BQ - \mathscr{D}_A\tilde{\gamma}^{AB}\mathscr{D}_BQ + \mathscr{D}^A\mathscr{D}_A\tilde{Q} - g^{AB}\mathscr{D}(\mathscr{D}_A\tilde{W}_B)Q = 0. \quad (4.30)$$

The dependent equation (3.11) gives us

$$\mathscr{D}^A\tilde{W}_A + \mathrm{Ad}(w)\tilde{w} + \omega(Q,\tilde{Q}) = 0. \quad (4.31)$$

After tedious algebra one can check that the following holds.

(i) Equation (4.31) is a consequence of (4.28)–(4.30).

(ii) By multiplying (4.27) with n and adding the divergence of (4.27) together with (4.23) one obtains minus the g product of (4.28) with w: This is clearly an nth-order Bianchi identity.

(iii) Taking the divergence of (4.22) and the covariant gradients of (4.23) and (4.24) and combining the result with the g product of (4.29) with B and the q product of (4.30) with ∂_AQ, one obtains Eq. (4.27): These equations are consequences of the two remaining, four-dimensional Bianchi identities.

Consequently, we can discard Eqs. (4.25), (4.27), and (4.31) and work with the system consisting of Eqs. (4.22)–(4.24) and (4.28)–(4.30).

V. THE CASE $\mathscr{G}$ = SU(2)

In this section we want to specialize the nth-order equations to the gauge group SU(2) and the non-Abelian zeroth-order solution described in Ref. 14. Of course, this non-Abelian solution will also be present for larger gauge groups as long as they contain an SU(2) subgroup. On the other hand, larger gauge groups may allow for a larger set of non-Albelian solutions.

Specializing the g and q inner products as shown in Ref. 14 we arrive at the desired set of equations. We split the YMH equations up into a longitudinal component, parallel in gauge space to the unit internal vector **n**, and the transversal components orthogonal to it. The internal vector **n** is chosen such that

$$\mathbf{Q} \doteq Q^a\mathbf{T}^a = q\mathbf{n} \doteq q(n^a\mathbf{T}^a).$$

In order to distinguish between Lie-algebra-valued quantities and their components relative to a given basis of the Lie algebra, from now on we will use boldface letters for the former and lightface letters for the latter. Lower case roman letters $a, b, c,...$ will be used as internal indices, running from 1–m, where m is the dimension of the gauge group; in our case $m = 3$. Our background solution is a nontrivial principal bundle over S^2. We choose to cover the minimal surface (t = const, $N = 0$) with two patches and choose different gauges, i.e., different $\mathbf{n}$, on each of the patches.[15] In the overlapping region a gauge transformation will relate both of the patches. We define the two patches R_n and R_s as follows:

$$R_n:\ 0\leqslant\theta<\pi/2+\delta,\quad \mathbf{n}_n=(1,0,0),$$

$$R_s:\ \pi/2-\delta\leqslant\theta\leqslant\pi,\quad \mathbf{n}_s=(-1,0,0).$$

The longitudinal part W^a_{pA} of the zeroth-order gauge potential W^a_A on each patch is given by

$$R_n:\ W^a_{npA}=(-\cos\theta+1)n^a_n,$$

$$R_s:\ W^a_{spA}=(\cos\theta+1)n^a_s.$$

We now introduce spherical coordinates in internal space:

$$\mathbf{T}^0\doteq\mathbf{T}^1,\quad \mathbf{T}^\pm\doteq(1/\sqrt{2})(\mathbf{T}^2\pm\mathbf{T}^3), \tag{5.1}$$

so that the commutation relations read as

$$[\mathbf{T}^0,\mathbf{T}^\pm]=\mp\iota\mathbf{T}^\pm,\quad [\mathbf{T}^+,\mathbf{T}^-]=-\iota\mathbf{T}^0.$$

The components of internal vectors in this new basis are related to the old ones via

$$\Phi^1\doteq\Phi^0,\quad \Phi^\pm\doteq(1/\sqrt{2})(\Phi^2\mp\iota\Phi^3).$$

We now proceed to specialize the nth-order equations to the non-Abelian zeroth-order solution. We begin with the Einstein equations (4.22)–(4.24):

$$\left[\frac{n(n+1)}{G^2}-4(\gamma g)^2\left(\frac{A}{r_0}\right)^2\right]\tilde{\gamma}_{AB}+\frac{2}{G}\left(\nabla_B\nabla_A\tilde{G}-\frac{1}{2}g_{AB}\Delta\tilde{G}\right)$$
$$=-8\gamma^2\left(\delta^K_{(A}\delta^L_{B)}-\frac{1}{2}g_{AB}g^{KL}\right)[-\iota q(\mathscr{D}^{\parallel}_K\tilde{Q}^+U^-_L$$
$$-\mathscr{D}^{\parallel}_K\tilde{Q}^-U^+_L)+q^2(U^+_L\tilde{W}^-_K+U^-_L\tilde{W}^+_K)], \tag{5.2}$$

$$(2/G)\Delta\tilde{G}=-[n(n-1)/G^2]\tilde{\kappa}, \tag{5.3}$$

$$-[2(n+2)/G^3]\tilde{G}+(n/G^2)\tilde{\kappa}+2(\gamma q)^2e^2q^2\tilde{\kappa}$$
$$=4(\gamma q)^2[e^{AB}\nabla_A\tilde{W}^0_B-\iota e^{AB}(U^+_A\tilde{W}^-_B$$
$$-U^-_A\tilde{W}^+_B)]+8\gamma^2q(A/r_0)^2\tilde{Q}^0. \tag{5.4}$$

We have introduced the following notation:

$$A\doteq(r_0/2)\sqrt{k(F^2-q^2)}, \tag{5.5}$$

$$\mathscr{D}^{\parallel}_K\doteq\nabla_A-\mathrm{Ad}(\mathbf{W}_{pK}), \tag{5.6}$$

and we adhere to the following convention:

$$\phi_{(A,B)}\doteq\tfrac{1}{2}(\phi_{AB}+\phi_{BA}). \tag{5.7}$$

Following the notation in Ref. 14 the full gauge potential is given by

$$W^a_K=W^a_{pK}+U^a_K=W_Kn^a+U^a_K. \tag{5.8}$$

The transversal components on the northern hemisphere are given in the new basis (5.1)

$$U^+_{n2}=(A/\sqrt{2})e^{\iota\phi},\quad U^+_{n3}=\iota(A/\sqrt{2})\sin\theta e^{\iota\phi},$$
$$U^-_{n2}=(A/\sqrt{2})e^{-\iota\phi},\quad U^-_{n3}=-\iota(A/\sqrt{2})\sin\theta e^{-\iota\phi}. \tag{5.9}$$

In the new basis we obtain

$$(\mathscr{D}^{\parallel}_K\phi)^\pm=(\nabla_K\mp\iota W_K)\phi^\pm. \tag{5.10}$$

From the symmetry condition (see, e.g., Ref. 16) for the gauge potential, one can easily show that

$$\mathscr{D}^{\parallel}_A\mathbf{U}_B=0. \tag{5.11}$$

Further, we have

$$U^+_AU^-_B+U^-_AU^+_B=\tfrac{1}{2}g_{AB}(U)^2, \tag{5.12}$$

with

$$(U)^2=g^{AB}(U^+_AU^-_B+U^-_AU^+_B). \tag{5.13}$$

The longitudinal component of the nth-order electric equation (4.28) is

$$\Delta\tilde{\omega}^0+[n(n+1)/G^2-2(A/r_0)^2]\tilde{\omega}^0$$
$$-2\iota g^{AB}[U^+_A\mathscr{D}^{\parallel}_B\tilde{\omega}^--U^-_A\mathscr{D}^{\parallel}_B\tilde{\omega}^+]=0. \tag{5.14}$$

The transversal components, after some algebra, read as

$$g^{AB}\mathscr{D}^{\parallel}_A\mathscr{D}^{\parallel}_B\tilde{\omega}^\pm+[n(n+1)/G^2-e^2q^2-(A/r_0)^2]\tilde{\omega}^\pm$$
$$\pm2\iota U^+_Ag^{AB}\partial_B\tilde{\omega}^0=0, \tag{5.15}$$

where we have used

$$g^{AB}U^+_B(U^+_A\tilde{\omega}^-+U^-_A\tilde{\omega}^+)=(A/r_0)^2\tilde{\omega}^\pm.$$

The longitudinal component of the Higgs equation (4.30) is

$$\Delta\tilde{Q}^0+[n(n+1)/G^2-kq^2]\tilde{Q}^0+q(A/r_0)^2\tilde{\kappa}$$
$$-2qg^{AB}(U^-_B\tilde{W}^+_A+U^+_B\tilde{W}^-_A)$$
$$-2\iota g^{AB}(U^+_A\mathscr{D}^{\parallel}_B\tilde{Q}^--U^-_A\mathscr{D}^{\parallel}_B\tilde{Q}^+)=0 \tag{5.16}$$

and the transversal components are

$$g^{AB}\mathscr{D}^{\parallel}_A\mathscr{D}^{\parallel}_B\tilde{Q}^\pm+[n(n+1)/G^2+(A/r_0)^2-e^2q^2]\tilde{Q}^\pm$$
$$\pm2\iota U^\pm_Ag^{AB}\partial_B\tilde{Q}^0\pm\iota(q/G)g^{AB}U^\pm_B\partial_A\tilde{G}$$
$$+2qg^{AB}U^\pm_B\tilde{W}^0_A\mp\iota qU^\pm_B\nabla_A\tilde{\gamma}^{AB}=0. \tag{5.17}$$

Here we have used the following zeroth-order result:

$$(\mathscr{D}_A\mathscr{D}_BQ)^a=-U^b_AU^b_Bqn^a=-\tfrac{1}{2}g_{AB}(U)^2qn^a$$

and the definition of $\tilde{\gamma}^{AB}$. Finally, for the magnetic equations we obtain

$$e^{AB}e^{KL}\nabla_B\nabla_K\tilde{W}^0_L-[n(n+1)/G^2-(A/r_0)^2]g^{AB}\tilde{W}^0_B$$
$$-2\iota e^{AB}e^{KL}(U^+_{(K}\mathscr{D}^{\parallel}_{B)}\tilde{W}^-_L-U^-_{(K}\mathscr{D}^{\parallel}_{B)}\tilde{W}^+_L)$$
$$+e^{AB}(e^2q^2/G)\partial_B\tilde{G}-\tfrac{1}{2}e^{AB}e^2q^2\partial_B\tilde{\kappa}$$
$$-e^2qg^{AB}(U^+_B\tilde{Q}^-+U^-_B\tilde{Q}^+)=0 \tag{5.18}$$

and

$$e^{AB}e^{KL}\mathscr{D}^{\parallel}_B\mathscr{D}^{\parallel}_K\tilde{W}^\pm_L-[n(n+1)/G^2-(A/r_0)^2-e^2q^2]g^{AB}\tilde{W}^\pm_{AB}\pm\iota e^{AB}e^{KL}(U^\pm_K\partial_B\tilde{W}^0_L+U^\pm_B\partial_K\tilde{W}^0_L)$$
$$+e^{AB}e^{KL}U^\pm_K(U^+_B\tilde{W}^-_L+U^-_B\tilde{W}^+_L)\pm\iota e^2q^2e^{AB}\tilde{W}^\pm_B-e^2q^2U^\pm_B(\tilde{\gamma}^{AB}-\tfrac{1}{2}g^{AB}\tilde{\kappa})$$
$$+2g^{AB}e^2q\tilde{Q}^0U^\pm_B\mp\iota e^2qg^{AB}\mathscr{D}^{\parallel}_B\tilde{Q}^\pm=0. \tag{5.19}$$

VI. HARMONIC EXPANSION

In order to solve Eqs. (5.2)–(5.4) and (5.14)–(5.19) we have to expand the nth-order fields according to their transformation properties under spin and isospin rotations. We write the correct expansions on R_n and refer to Refs. 17 and 13 for more details. The expansions on R_s are analogous. We will not use them in our calculations.

$$\tilde{\kappa} = \sum_{l,m} \tilde{\kappa}_{l,m} Y_{l,m} \tag{6.1}$$

$$\tilde{\gamma}_{ab} = \sum_{l,m} \Big[\tilde{\gamma}_{l,m} \Big(\nabla_A \nabla_B - \frac{1}{2} g_{AB} \Delta \Big) Y_{l,m} + \tilde{\kappa}_{l,m} \frac{1}{2} (e_A{}^K \nabla_K \nabla_B + e_B{}^K \nabla_K \nabla_A) Y_{l,m} \Big], \tag{6.2}$$

$$\tilde{w}^{\pm} = \sum_{l,m} \tilde{u}^{\pm}_{l,m} Y^{\pm}_{l,m}, \tag{6.3}$$

$$\tilde{w}^0 = \sum_{l,m} \tilde{w}^0_{l,m} Y_{l,m}, \tag{6.4}$$

$$\tilde{Q}^{\pm} = \sum_{l,m} \tilde{Q}^{\pm}_{l,m} Y^{\pm}_{l,m}, \tag{6.5}$$

$$\tilde{Q}^0 = \sum_{l,m} \tilde{Q}^0_{l,m} Y_{l,m}, \tag{6.6}$$

$$\tilde{W}^{\pm}_A = \sum_{l,m} [\tilde{W}^{\pm+}_{l,m} D^{\|}_A Y^{\pm}_{l,m} + \tilde{W}^{\pm-}_{l,m} e_A{}^K D^{\|}_K Y^{\pm}_{l,m}], \tag{6.7}$$

$$\tilde{W}^0_A = \sum_{l,m} [\tilde{W}^{0+}_{l,m} \nabla_A Y_{l,m} + \tilde{W}^{0-}_{l,m} e_A{}^K \nabla_K Y_{l,m}], \tag{6.8}$$

where the second superscript on the expansion coefficients of $\tilde{W}^{\pm}_A$ and $\tilde{W}^0_A$ denotes even (+) or odd (−) parity character. The $Y_{l,m}$ are the usual spherical harmonics, which are a special case of the monopole harmonics $Y_{q,l,m}$ for $q = 0$.[18] Monopole harmonics are defined separately on each patch:

$$R_n: \quad Y_{n;q,l,m}(\theta,\phi) = (1/4\pi) e^{\iota(m+q)\phi} P_{q,l,m}(\cos\theta), \tag{6.9}$$

$$R_s: \quad Y_{s;q,l,m}(\theta,\phi) = (1/\sqrt{4\pi}) e^{\iota(m-q)\phi} P_{q,l,m}(\cos\theta), \tag{6.10}$$

with

$$P_{q,l,m}(\mu) = \left[\frac{(l-m)!}{(l+m)!} (2l+1)! \frac{1}{(l-b)!(l+b)!} \right]^{1/2} \times \frac{1}{2^l} (1+\mu)^{(m-b)/2} (1-\mu)^{(m+b)/2} \times \left(-\frac{d}{d\mu} \right)^{l+m} (1+\mu)^{l+b} (1-\mu)^{l-b} \tag{6.11}$$

on both patches. On each patch the monopole harmonics satisfy the relation (see, e.g. Refs. 19 and 20)

$$Y_{q,l,m} Y_{q',l',m'} = \sum_{l''} (-1)^{l+l'+l''+m''+q''} \times \left[\frac{(2l''+1)(2l''+2)(2l''+3)}{4\pi} \right]^{1/2} \times \begin{pmatrix} l & l' & l'' \\ m & m' & m'' \end{pmatrix} \begin{pmatrix} l & l' & l'' \\ q & q' & q'' \end{pmatrix} Y_{-q'',l'',-m''}, \tag{6.12}$$

where

$$\begin{pmatrix} l & l' & l'' \\ a & b & c \end{pmatrix}$$

are the Wigner $3j$ symbols.

The functions $Y^{\pm}_{l,m}$ are closely related to the monopole harmonics[13]:

$$Y^{\pm}_{l,m} = Y_{\pm q,l,m}.$$

For our background solution, $q = 1$.

In order for the fields to be real, respectively, anti-Hermitian, the expansion coefficients have to satisfy the following conditions:

$$\Phi^{\pm}_{l,m} = \iota^{l+m} \Phi^{+}_{R;l,m}, \quad \Phi^{-}_{l,m} = \iota^{-1+m} \Phi^{-}_{R;l,m}, \tag{6.13}$$

$$(\Phi^{+}_{R;l,m})^* = \Phi^{+}_{R;l,m}, \quad \Phi^{+}_{R;l,m} = (-1)^{1+m} \Phi^{-}_{R;l,-m}, \tag{6.14}$$

$$\Phi^0_{l,m} = \Phi^0_{l,-m}(-1)^m, \quad (\Phi^0_{lm})^* = \Phi^0_{lm}, \tag{6.15}$$

where the asterisk means complex conjugation. This follows directly from the formula

$$(Y_{q,l,m})^* = (-1)^{q+m} Y_{-q,l,-m}. \tag{6.16}$$

VII. ALGEBRAIZATION OF THE nth-ORDER EQUATIONS

Let us start with the longitudinal electric equation (5.14) and concentrate on the third term. The third term is the only one that is not straightforward to compute. It is helpful to start with the following expression:

$$g^{AB}(U^+_A \mathscr{D}^{\|}_B Y^-_{l,m} - U^-_A \mathscr{D}^{\|}_B Y^+_{l,m}) = \frac{A}{\sqrt{2} r_0^2} \Big[e^{\iota\phi} \mathscr{D}^{\|}_2 Y^-_{l,m} + \frac{\iota}{\sin\theta} e^{\iota\phi} \mathscr{D}^{\|}_3 Y^-_{l,m} + e^{-\iota\phi} \mathscr{D}^{\|}_2 Y^+_{l,m} + \frac{\iota}{\sin\theta} e^{-\iota\phi} \mathscr{D}^{\|}_3 Y^+_{l,m} \Big], \tag{7.1}$$

where we have used (5.9). Now we insert the following relations, which follow directly from the definition of the $Y^{\pm}_{l,m}$[13]:

$$\mathscr{D}^{\|}_2 Y^{\pm}_{l,m} = \tfrac{1}{2} [e^{-\iota\phi} \sigma_{m+1} Y^{\pm}_{l,m+1} - e^{\iota\phi} \sigma_m Y^{\pm}_{l,m-1}], \tag{7.2}$$

$$\mathscr{D}^{\|}_3 Y^{\pm}_{l,m} = -\iota \sin\theta \, [\tfrac{1}{2} e^{-\iota\phi} \cos\theta \sigma_{m+1} Y^{\pm}_{l,m+1} + \tfrac{1}{2} e^{\iota\phi} \cos\theta \sigma_m Y^{\pm}_{l,m-1} - m \sin\theta \, Y^{\pm}_{l,m}], \tag{7.3}$$

with

$$\sigma_m \doteq \sqrt{(l+m)(l-m+1)}$$

and we obtain

$$g^{AB}(U^+_A \mathscr{D}^{\|}_B Y^-_{l,m} - U^-_A \mathscr{D}^{\|}_B Y^+_{l,m}) = (A/r_0^2\sqrt{2}) [\tfrac{1}{2}\sigma_{m+1}(1+\cos\theta) Y^-_{l,m+1} - \tfrac{1}{2}\sigma_m e^{2\iota\phi}(1-\cos\theta) y^-_{l,m-1} - \tfrac{1}{2}\sigma_{m+1} e^{-2\iota\phi}(1-\cos\theta) Y^+_{l,m+1} + \tfrac{1}{2}\sigma_m (1+\cos\theta) Y^+_{l,m-1} - e^{\iota\phi} m \sin\theta Y^-_{l,m} - e^{\iota\phi} m \sin\theta Y^+_{l,m}]. \tag{7.4}$$

Because

$$Y^{\pm}_{1,\pm} = \frac{1}{\sqrt{4\pi}}\sqrt{\frac{3}{4}}\,e^{\pm 2\iota\phi}(1-\cos\theta),$$

$$Y^{\pm}_{1,\mp} = \frac{1}{\sqrt{4\pi}}\sqrt{\frac{3}{4}}(1+\cos\theta), \tag{7.5}$$

$$Y^{\pm}_{1,0} = \mp\sqrt{\frac{3}{8\pi}}\,e^{\iota\phi}\sin\theta,$$

this means that

$$\begin{aligned} &g^{AB}(U^+_A\mathscr{D}^{\parallel}_B Y^-_{l,m} - U^-_A\mathscr{D}^{\parallel}_B Y^+_{l,m}) \\ &\quad= (A/\sqrt{2}r_0^2)\sqrt{4\pi/3}\,[\sigma_{m-1}Y^+_{1,-1}Y^-_{l,m+1} \\ &\qquad -\sigma_m Y^+_{1,1}Y^-_{l,m-1} - \sigma_{m+1}Y^-_{1,-1}Y^+_{l,m+1} \\ &\qquad +\sigma_m Y^-_{1,1}Y^+_{l,m-1} \\ &\qquad +\sqrt{2}mY^+_{1,0}Y^-_{l,m} - \sqrt{2}mY^-_{1,0}Y^+_{l,m}]\,. \end{aligned} \tag{7.6}$$

We now go back to Eq. (5.14) and insert the expansions (6.4) and (6.3). Using result (7.6), the addition formula (6.13), and the orthogonality of the spherical harmonics, we obtain

$$\begin{aligned} &\left(\frac{n(n+1)}{G^2} - \frac{l(l+1)}{r_0^2} - 2\left(\frac{A}{r_0}\right)^2\right)\tilde{\omega}^0_{l,m}Y_{l,m} \\ &\quad -2\iota\frac{A}{r_0^2\sqrt{2}}\sqrt{l(l+1)}\,(\tilde{\omega}^+_{l,m} + \tilde{\omega}^-_{l,m}) = 0\,. \end{aligned} \tag{7.7}$$

Because of (6.13) and (6.14) the rhs of Eq. (7.7) is equal to

$$\iota^m(\tilde{\omega}^-_{R;l,m}(-1)^{1+m} - \tilde{\omega}^-_{R;l,m})\,.$$

Without loss of generality we can set $m=0$ because the background is spherically symmetric up to isospin transformations: This simplifies the calculations. Hence, for the longitudinal electric equation, we finally obtain

$$\begin{aligned} &\left(\frac{n(n+1)}{G^2} - \frac{l(l+1)}{r_0^2} - 2\left(\frac{A}{r_0}\right)^2\right) \\ &\quad\times\tilde{\omega}^0_{l,0} \pm 2\frac{A}{r_0^2}\sqrt{2}C\tilde{\omega}^{\pm}_{R;l,o} = 0\,. \end{aligned} \tag{7.8}$$

For the transversal equation the calculations are very similar. In addition, one needs the following formulas:

$$g^{AB}\mathscr{D}^{\parallel}_A\mathscr{D}^{\parallel}_B Y^{\pm}_{l,m} = (1/r_0^2)[-l(l+1)+1]Y^{\pm}_{l,m}\,, \tag{7.9}$$

which follows directly from the definition of the $Y_{q,l,m}$ as eigenfunctions of the generalized angular momentum operators, and

$$e^2q^2 = (1-A^2)/r_0\,, \tag{7.10}$$

which is a property of our background. With these results we obtain

$$(n(n+1)/G^2 - l(l+1)/r_0^2)\tilde{\omega}^{\pm}_{R;l,0} \pm (A/r_0^2)\sqrt{2}C\tilde{\omega}^0_{l,0} = 0, \tag{7.11}$$

where $C \doteq \sqrt{l(l+1)}$. From now on we will omit the $m=0$ index. The following relations can be derived from the defining Eqs. (6.9), (6.11), and (5.9):

$$U^{\pm}_K Y_{l0} = (A/\sqrt{2}C)[\mp\mathscr{D}^{\parallel}_K Y^{\pm}_{l0} + \iota e_K{}^L\mathscr{D}^{\parallel}_L Y^{\pm}_{l0}]\,,$$

$$Y^+_{l0}U^-_K + Y^-_{l0}U^+_K = \iota(A\sqrt{2}/C)e_K{}^L\nabla_L Y_{l0}\,.$$

Performing similar calculations as above we obtain, for the scalar equations,

$$\begin{aligned} &\left(\frac{n(n+1)}{G^2} - \frac{l(l+1)}{r_0^2} + 2\left(\frac{A}{r_0}\right)^2\right)\tilde{Q}^{\pm}_{R;l} \\ &\quad \pm C\Big[2q\tilde{W}^{0-}_l + \frac{q}{G}\tilde{G}_l - \frac{1}{2}\frac{q}{r_0} \\ &\quad \times[-l(l+1)+2]\tilde{\gamma}_l + 2\tilde{Q}_l\Big]\frac{A}{r_0^2\sqrt{2}} = 0\,, \end{aligned} \tag{7.12}$$

$$(2q\tilde{W}^{0+}_l + \tfrac{1}{2}(q/r_0^2)[-l(l+1)+2]\tilde{k}_l)C(A/r_0^2\sqrt{2}) = 0\,, \tag{7.13}$$

and

$$\begin{aligned} &\left(\frac{n(n+1)}{G^2} - \frac{l(l+1)}{r_0^2} - kq^2\right)\tilde{Q}^0_l + \left(\frac{A}{r_0}\right)^2 q\tilde{\kappa}_l \\ &\quad + 2\frac{\sqrt{2}A}{r_0}C(-q\tilde{W}^{+}_{R;l}{}^{-} + \tilde{Q}^+_{R;l}) = 0\,. \end{aligned} \tag{7.14}$$

Equations (7.12) and (7.13) result from the transversal Higgs equation (5.17) and Eq. (7.14) results from the longitudinal part (5.16). For $n=1$ the term proportional to $\tilde{G}_l$ does not appear in Eq. (7.12) because it arises from a term containing a derivative of $\tilde{G}$ and for $n=1$ Eq. (4.23) tells us that $\tilde{G}$ is constant. The transversal magnetic equation (5.19) leads, after inserting the expansions and extensively using the abovementioned formulas to the following system:

$$(n(n+1)/G^2 - e^2q^2)\tilde{W}^{+}_{R;l}{}^{+} - \tfrac{1}{2}(A/\sqrt{2}r_0^2)Ce^2q^2\tilde{k}_l = 0\,, \tag{7.15}$$

$$\begin{aligned} &\left(e^2q^2 - \left(\frac{A}{r_0}\right)^2\right)\tilde{W}^{+}_{R;l}{}^{-} + \frac{A}{\sqrt{2}r_0^2}C\tilde{W}^{0-}_{R;l} - \frac{1}{2}e^2q^2\frac{A}{\sqrt{2}r_0^2}C\tilde{\gamma}_l \\ &\quad -\frac{1}{2}e^2q^2\frac{A}{\sqrt{2}C}\tilde{\kappa}_l - 2e^2q\frac{A}{\sqrt{2}C}\tilde{Q}^0_l + e^2q\tilde{Q}^{\pm}_{R;l} = 0\,, \end{aligned} \tag{7.16}$$

$$\begin{aligned} &\left(\frac{n(n+1)}{G^2} - \frac{l(l+1)}{r_0^2}\right)\tilde{W}^{+}_{R;l}{}^{-} + \frac{A\sqrt{2}}{r_0^2}C\tilde{W}^{0-}_l \\ &\quad -\frac{1}{2}e^2q^2\frac{A}{\sqrt{2}C}\tilde{\kappa}_l - 2e^2\frac{A}{\sqrt{2}C}\tilde{Q}^0_l \\ &\quad +\frac{1}{2}e^2q^2\frac{A}{\sqrt{2}r_0^2}C\tilde{\gamma}_l = 0\,, \end{aligned} \tag{7.17}$$

$$(A/\sqrt{2}r_0^2)C(\tilde{W}^{0+}_l - \tfrac{1}{2}e^2q^2\tilde{k}_l) + (A/r_0)^2\tilde{W}^{+}_{R;l}{}^{+} = 0\,, \tag{7.18}$$

and the longitudinal component (5.18),

$$\begin{aligned} &(n(n+1)/G^2 - (A/r_0)^2)\tilde{W}^{0+}_l \\ &\quad -2(A/\sqrt{2}r_0^2C)\tilde{W}^{+}_{R;l}{}^{+} = 0 \end{aligned} \tag{7.19}$$

and

$$\left(\frac{n(n+1)}{G^2}-\frac{l(l+1)}{r_0}-\left(\frac{A}{r_0}\right)^2\right)\widetilde{W}_l^{0-}$$
$$-e^2q^2\frac{\widetilde{G}_l}{G}+\frac{1}{2}e^2q^2\tilde{\kappa}_l$$
$$+4\frac{A}{\sqrt{2}r_0^2C}\left(l(l+1)-\frac{1}{2}\right)\widetilde{W}_{R;l}^{+-}$$
$$+2e^2q\frac{A}{\sqrt{2}C}\widetilde{Q}_{R;l}^{+}=0. \quad (7.20)$$

Again, the term porportional to $\widetilde{G}_l$ does not appear in (7.20) if $n=1$.

The symmetric traceless part (5.2) of Einstein's equations yields

$$\left(\frac{n(n+1)}{G^2}-4(\gamma q)^2\left(\frac{A}{r_0}\right)^2\right)\tilde{\gamma}_l+\frac{2}{G}\widetilde{G}_l$$
$$=-8\gamma^2q\frac{\sqrt{2}A}{C}(q\widetilde{W}_{R;l}^{+-}+\widetilde{Q}_{R;l}^{+}) \quad (7.21)$$

and

$$(n(n+1)/G^2-4(\gamma q)^2(A/r_0)^2)\tilde{k}_l$$
$$=8(\gamma q)^2(A\sqrt{2}/C)\widetilde{W}_{R;l}^{++}. \quad (7.22)$$

The trace part (5.3) immediately leads to

$$[l(l+1)/r_0^2]\widetilde{G}_l=\tfrac{1}{2}[n(n-1)/G]\tilde{k}_l. \quad (7.23)$$

Finally, (5.4) yields

$$-\frac{2(n+2)}{G^2}\frac{\widetilde{G}_l}{G}+\frac{n}{G^2}\tilde{\kappa}_l+2(\gamma q)^2e^2q^2\tilde{\kappa}_l-8\gamma^2\left(\frac{A}{r_0}\right)^2q\widetilde{Q}_l^0$$
$$=4(\gamma q)^2\left[\frac{l(l+1)}{r_0}\widetilde{W}_l^{0-}+\frac{A\sqrt{2}}{r_0^2}C\widetilde{W}_{R;l}^{+-}\right]. \quad (7.24)$$

The zeroth-order quantities G and r_0 are given by[14]

$$(\gamma/G)^2=e^2(\gamma q)^4-(k/4)[(\gamma F)^2-(\gamma q)^2]^2-\gamma^2\Lambda \quad (7.25)$$

and

$$1/r_0^2=1/G^2+(k/2)(\gamma F)^2(F^2-q^2)+2\Lambda. \quad (7.26)$$

VIII. DISCUSSION

To begin, let us write the range of the parameters allowed by the non-Abelian zeroth-order solution in a unified way for the three cases discussed in Ref. 14, namely case (i) $k=4e^2$, case (ii) $k<4e^2$, and case (iii) $k>4e^2$. The cosmological constant can be written in each case as

$$4\gamma^2\Lambda/k=e^2/k-s[(\gamma F)^2-\tfrac{1}{2}]^2, \quad (8.1)$$

where the parameter s has the following range:

case (i) $k=4e^2$, $s=4e^2/k=1$,

case (ii) $k<4e^2$, $1\leqslant s<4e^2/k$,

case (iii) $k<4e^2$, $1\geqslant s>4e^2/k$.

In case (i) $(\gamma q)^2$ is a free parameter within the limits $\frac{1}{2}\leqslant(\gamma q)^2<(\gamma F)^2$. The lower limit corresponds to $1/G^2=0$ and the upper limit corresponds to the Abelian case $Q=F$. The scalar energy density at the horizon decreases from $(e^2/2\gamma^4)[(\gamma F)^2-\frac{1}{4}]^2$ to zero and the YM energy density increases from $e^2/8\gamma^4$ to $(e^2/2\gamma^4)(\gamma F)^4$. The radius of the hole r_0 does not depend on $(\gamma q)^2$; the geometry changes through G. In cases (ii) and (iii) $(\gamma q)^2$ is fixed by

$$(\gamma q)^2=\tfrac{1}{2}+[(\gamma F)^2-\tfrac{1}{2}]\sqrt{k(1-s)/(k-4e^2)}. \quad (8.2)$$

In case (ii) the lower limit corresponds to $Q=F$ and the upper limit corresponds to $1/G^2=0$. The opposite is true for the case (iii).

We now want to find out when the system of equations (7.8) and (7.11)–(7.24) possesses nonzero solutions for this allowed range of zeroth-order values. One remarks first that the system is simplified by the fact that the electric fields $\tilde{w}^{\pm}$, $\tilde{w}^0$ do not couple to the rest of the fields. The condition for a nonzero solution is

$$\left[\frac{n(n+1)}{G^2}-\frac{l(l+1)}{r_0^2}+2\left(\frac{A}{r_0}\right)^2\right]\left[\frac{n(n+1)}{G^2}-\left(\frac{A}{r_0}\right)^2\right]$$
$$-4\frac{n(n+1)}{G^2}\left(\frac{A}{r_0}\right)^2=0, \quad (8.3)$$

where one should remember that it is valid for $n>1$ only if all lower-order solutions vanish. If, e.g., Eq. (8.3) can be satisfied for $n=1$ by choosing appropriate parameters, then the second-order electric equations will in general differ from Eqs. (7.8) and (7.11) with $n=2$ by terms linear and quadratic in first-order quantities.

Therefore, the correct way to proceed is to start by analyzing the system for $n=1$. Once this is solved, one knows the conditions for the second-order system to be valid etc. However, we remark that since the general nth-order equations (including lower-order contributions) are also symmetric under general rotations, there will be no mixing of different l terms for the nth-order quantities and the lower-order contributions will combine into terms with given l. Hence, the general nth-order equations will split for every l into the special nth-order equations we have derived plus nonhomogeneous terms containing the lower-order contributions. Now, if one considers only nonvanishing lower-order contributions with $l=0$, the resulting term in the general nth-order equations will be a pure $l=0$ term, thus leaving the $l>0$ components untouched. In other words, the existence of a nonzero $l=0$ first-order solution will still lead to second-order equations given by (7.8) and (7.11)–(7.24) with $n=2$ for $l>0$, and so on for higher orders. For $l>0$ lower-order contributions the picture gets more complicated.

Before proceeding, let us remark that the odd-parity fields decouple from the even-parity fields. This must be so because parity commutes (in a more subtle sense than usual, i.e., one must be careful to choose the correct gauge in the corresponding patch) with generalized rotations.

A. Odd parity

For $l=0$, 1 the equations describing the odd-parity fields only have the trivial solution. For $l>2$ we have to solve an overdeterminate system of linear equations. We first consider the system consisting of Eqs. (7.15), (7.19), and

(7.22). Setting the determinant to zero yields the characteristic equation

$$\frac{n(n+1)}{G^2}\left[\left(\frac{n(n+1)}{G^2}\right)^2-\frac{n(n+1)}{G^2}\left(\frac{1}{r_0^2}+4(\gamma q)^2\left(\frac{A}{r_0}\right)^2\right)+\left(\frac{A}{r_0}\right)^2\left(4(\gamma q)^2\left(\frac{A}{r_0}\right)^2+\frac{1}{r_0^2}-\left(\frac{A}{r_0}\right)^2\right)\right]=0,$$

where we have used $e^2q^2=(1-A^2)/r_0^2$. The remarkable fact is that the determinant is independent of l. The above equation has three solutions:

(i) $1/G^2=0$,

(ii) $n(n+1)/G^2=(A/r_0)^2$,

(iii) $n(n+1)/G^2=\frac{1}{r_0^2}+4(\gamma q)^2\left(\frac{A}{r_0}\right)^2-\left(\frac{A}{r_0}\right)^2$.

It is easy to check that for case (i) Eq. (7.18) is also satisfied, but (7.13) requires $l(l+1)=4$ for $q\neq 0$, i.e., there is no nonzero solution.

Case (ii) leads to a solution where $\widetilde{W}_l^{0+}$ is arbitrary and $\widetilde{W}_l^{++}=0=\tilde{k}_l$. However, only $\widetilde{W}_l^{0+}=0$ satisfies Eq. (7.13) with $q\neq 0$. Hence, there is only the trivial solution in this case.

Case (iii) leads to a family of solutions with $\widetilde{W}_l^{++}$ arbitrary:

$$\tilde{k}_l=[8(\gamma q)^2\sqrt{2A}/Ce^2q^2]\,\widetilde{W}_l^{++},$$

$$\widetilde{W}_l^{0+}=\frac{A\sqrt{2}}{C}\frac{1}{1-2A^2(1-2(\gamma q)^2)}\widetilde{W}_l^{++},$$

However, Eq. (7.18) restricts the solutions to either $(\gamma q)^2=\frac{1}{2}$, which means $1/G^2=0$, or

$$4(\gamma q)^4-(4e^2/k-1)(\gamma q)^4-(\gamma q)^2(4(\gamma F)^2+1)-(\gamma F)^4+(\gamma F)^2-4\gamma^2\Lambda/k=0. \quad (8.4)$$

Using Eq. (53) of Ref. 4 condition (8.4) is equivalent to

$$(\gamma q)^2=(\gamma F)^2+e^2/k.$$

Hence, (8.4) cannot be satisfied for any q in the range $0<q<F$.

Hence, we conclude that there are no odd-parity solutions for $q\neq 0$ for any $n>1$. This implies that for $q\neq 0$ any global solution to the EYMH system satisfying our assumptions will have no odd-parity components near the horizon.

B. Even parity; $n=1$, $l=0$

For $l=0$, $n=1$ the system reduces to

$$(2/G^2-kq^2)\widetilde{Q}_0^0+(A/r_0)^2q\tilde{\kappa}_0=0, \quad (8.5)$$

$$(2/G^2-2(A/r_0)^2)\widetilde{w}_0^0=0, \quad (8.6)$$

$$-\tfrac{1}{2}(A/\sqrt{2}e^2q^2\tilde{\kappa}_0-2e^2q(A/\sqrt{2})\widetilde{Q}_0^0=0, \quad (8.7)$$

$$-(6/G^3)\widetilde{G}+(1/G^2)\tilde{\kappa}_0+2\gamma^2e^2q^4\tilde{\kappa}_0-8\gamma^2(A/r_0)^2q\widetilde{Q}_0^0=0. \quad (8.8)$$

From Eq. (8.6) it follows that $\widetilde{w}_0^0$ will vanish unless (8.3) with $n=1$, $l=0$ is satisfied. For the rest of the system the condition for the existence of a nontrivial solution assuming $q\neq 0$, F and $1/G^2\neq 0$ is

$$2/G^2-4(A/r_0)^2-kq^2=0. \quad (8.9)$$

If $q=0$ (Abelian solution), then $\tilde{Q}^0$ and $\kappa=6(\tilde{G}_0/G)$. The cases $q=0$, F lead to Abelian zeroth-order solutions and have been treated elsewhere[8]. Using (7.25), (8.9) results in a quadratic equation for $(\gamma q)^2$:

$$(\gamma q)^4(4e^2-k)+2k(\gamma q)^2(\gamma F)^2-k(\gamma F)^4-4\Lambda\gamma^2-2k(\gamma F)^2=0.$$

Again using Eq. (53) of Ref. 14 we obtain

$$(4e^2-k)(\gamma q)^2-2k(\gamma F)^4+2k(\gamma F)^2[(\gamma q)^2-\tfrac{1}{2}]-8\Lambda\gamma^2=0. \quad (8.10)$$

For $\kappa=4e^2$ Eq. (8.10) only has one solution:

$$(\gamma q)^2=\tfrac{3}{2}.$$

As $(\gamma q)^2<(\gamma F)^2$, this implies that we must choose $(\gamma F)^2>\frac{3}{2}$ and $4\gamma^2\Lambda<-3e^2$ (see Ref. 14). For $\kappa<4e^2$ Eq. (8.10), together with the inequality (40) in Ref. 14 [notice that the inequalities (38) and (40) in Ref. 14 should be interchanged] tells us that $(\gamma q)^2$ is constrained to the values

$$\tfrac{1}{2}\leqslant(\gamma q)^2<\tfrac{3}{2}.$$

For $k>4e^2$ the analogous analysis tells us that

$$(\gamma q)^2>\tfrac{3}{2}.$$

For $q=0=F$ condition (8.9) is clearly equivalent to condition (8.3) for $n=1$, $l=0$.

C. Even parity; $n=1$, $l>0$

Because $\widetilde{G}_l=0$ for $l>0$, as a result of Eq. (7.23), the first-order system of equations for even solutions, (7.12), (7.14), (7.17), (7.20), (7.21), and (7.24), looks like an eigenvalue problem with the eigenvalue $n(n+1)/G^2$. In order to know for which values of G this is true, we have to solve the characteristic equation. We have performed this calculation for the case $k=4e^2$. For this case we have

$$A=\sqrt{1-(q/F)^2},$$

$$2/G^2=(4/r_0^2)((\gamma q)^2-\tfrac{1}{2}),$$

$$4\gamma^2\Lambda=e^2-4e^2[(\gamma q)^2-\tfrac{1}{2}]^2,$$

where $\frac{1}{2}\leqslant(\gamma q)^2<(\gamma F)^2$. For the uninteresting limiting case $q=F$ (Abelian background), we obtain a vanishing determinant in the following cases:

$$(\gamma q)^2=\tfrac{1}{2}, \quad (8.11)$$

$$(\gamma q)^2=[n(n+1)+C^2]/[2n(n+1)+4], \quad (8.12)$$

$$(\gamma q)^2=[n(n+1)+C^2]/2n(n+1), \quad (8.13)$$

$$(\gamma q)^2=[n(n+1)+C^2+4]/2n(n+1), \quad (8.14)$$

The first solution corresponds to the special case $1/G^2=0$ with $4\gamma^2\Lambda=e^2$ and $1/r_0^2=2\Lambda$. Restricting Λ to zero, we only found the following possibility if $Q\neq F$:

$$l=1,\quad (\gamma F)^2=1,\quad (\gamma q)^2=0.571.$$

The other cases are under investigation.

IX. SUMMARY AND CONCLUSIONS

We have derived the general first-order and special nth-order system of equations for extreme wormhole solutions in the EYMH system. We then specialized the system for the non-Abelian zeroth-order solution given in Ref. 14 and found that for $q\neq 0$ there are no odd-parity contributions. Solutions with $l=0$ exist in the case $k=4e^2$ only for $(\gamma q)^2=\frac{3}{2}$, exist in the case $k<4e^2$ only for $(\gamma q)^2<\frac{3}{2}$, and exist in the case $k>4e^2$ only if $(\gamma q)^2>\frac{3}{2}$. If we restrict in case (i) the cosmological constant Λ to zero, then there is only one $l>0$ solution $q\neq F$, namely,

$$l=1, \quad (\gamma F)^2=1 \quad (\gamma q)^2=0.571$$

and four with $q=F$.

This result strongly suggest that at least for $k=4e^2$, $\Lambda=0$, asymptotically flat non-Abelian solutions will only be possible for $(\gamma F)^2=1$, $(\gamma q)^2=0.571$. Because the zeroth-order solutions for $k>4e^2$ and $k<4e^2$ have no free parameters (once one has chosen Λ), as opposed to the case $k=4e^2$ with one free parameter $(\gamma q)^2$, it seems less likely to find nontrivial nth-order solutions in those cases. Our results seem to indicate that at least in the extreme case, non-Abelian static wormhole solutions of the EYMH system are only possible, if at all, by fine tuning the free parameters of the model.

ACKNOWLEDGMENTS

It is a pleasure to thank the members of the relativity group at the University of Wisconsin—Milwaukee and especially Professor J. Friedman for their warm hospitality and for many helpful discussions. I am also very much indebted to P. Hajicek for criticism and help at early stages of this work.

This work was supported by the Schweizerischer Nationalfonds.

[1]P. Hajicek, Nucl. Phys. B **185**, 254 (1981).
[2]G. Gibbons, in *Supersymmetry, Supergravity and Related Topics*, edited by F. del Aguila, J. A. Azcarraga, and L. E. Ibañez (World Scientific, Singapore, 1984).
[3]A. Salam and J. Strathdee, Phys. Lett. **61B**, 375 (1976).
[4]P. Hajicek, Phys. Lett. **106B**, 77 (1981).
[5]G. 't Hooft, Phys. Lett. **198B**, 61 (1987).
[6]P. van Nieuwenhuizen, D. Wilkinson, and M. J. Perry, Phys. Rev. D **13**, 778 (1976).
[7]P. Hajicek, Phys. Rev. D **26**, 3384 (1982); Phys. Rev. 3396 (1982).
[8]F. Degen, Gen. Rel. Grav. **19**, 739 (1987).
[9]P. Hajicek, Proc. R. Soc. London Ser. A **386**, 223 (1983).
[10]W. Israel, Commun. Math. Phys. **8**, 245 (1968).
[11]J. M. Arms, J. Math. Phys. **20**, 443 (1979).
[12]R. Bartnick and J. McKinnon, Phys. Rev. Lett. **61**, 141 (1988).
[13]F. Degen, Ph. D. thesis, Universität Bern (unpublished).
[14]P. Hajicek, J. Phys. A **16**, 1191 (1983).
[15]T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845 (1975).
[16]N. S. Manton and P. Forgacs, Commun. Math. Phys. **72**, 15 (1980).
[17]T. Regge and J. A. Wheeler, Phys. Rev. **108**, 1063 (1957).
[18]T. T. Wu and C. N. Yang, Nucl. Phys. B **107**, 365 (1976).
[19]T. T. Wu and C. N. Yang, Phys. Rev. D **16**, 1018 (1977).
[20]T. Dray, J. Math. Phys. **27**, 781 (1986).

Einstein–Maxwell equations and the groups of homothetic motion

Abbas M. Faridi
Department of Physics, California State University, Dominguez Hills Carson, California 90747

(Received 22 March 1989; accepted for publication 27 September 1989)

The Einstein–Maxwell field equations for a source-free, non-null electromagnetic field are studied under the assumption of admitting a nontrivial homothetic conformal motion, generating a homothetic bivector which is also non-null. It is shown that a space-time, whether vacuum or not, cannot admit a non-null homothetic vector field as a geodesic tangent. It is also shown that if a common principle null direction of the electromagnetic and the homothetic bivectors is geodesic and shear free, then the space-time must be algebraically special. Furthermore, it is found that if the electromagnetic and the homothetic bivectors have common principal null directions, then the vector field generating the homothetic bivector cannot be hypersurface orthogonal, unless it is a Killing vector field. Moreover, if these common principle null directions are also geodesics, then there exists no solution to the combined Einstein–Maxwell equations, unless, the non-null homothetic vector field is a Killing vector field. Finally, an example of a space-time admitting a non-null, nontrivial homothetic vector field generating a homothetic bivector which is also non-null is given.

I. INTRODUCTION

The importance of groups of motions, generated by the Killing vector fields of a space-time and their relation to the conservation laws of energy, momentum, and angular momentum is well known.[1–3] The properties of certain classes of vacuum space-times admitting a Killing vector field that gives rise to either a null or a non-null Killing bivector field have also been investigated.[4,5]

Collineations other than groups of motion have been studied to some extent, and in particular, it has been shown that for space-times with zero Ricci tensor, the more familiar symmetries such as motions, and conformal motions are subcases of a more general symmetry requirement known as curvature collineations.[6,7]

The significance of a homothetic conformal motion in general relativity is not yet fully understood; although much has been said about its group theoretic properties by various authors. For example, a topological description of conformal Killing vector fields on time-like two-surfaces is given by Plessis,[8] and self-similar spherically symmetric space-times are analyzed by Cahill and Taub.[9] Taub,[10] has emphasized the physical significance of self-similarity in general relativity in connection with plane-symmetric space-times, while Godfrey[11] has constructed all homothetic Weyl space-times. The role of conformal groups of transformations in relation to the relativistic wave equations and in Einstein–Maxwell theory are also discussed in Refs. 12–14.

In this paper we are concerned with source-free Einstein–Maxwell field equations

$$R_{\mu\nu} = f_{\mu\sigma} f^{\sigma}_{\nu} - \tfrac{1}{4} g_{\mu\nu} f_{\alpha\beta} f^{\alpha\beta}, \tag{1.1}$$

$$f^{\mu\nu}{}_{;\nu} = 0, \tag{1.2}$$

$${}^{*}f^{\mu\nu}{}_{;\nu} = 0, \tag{1.3}$$

with a non-null electromagnetic field, admitting a homothetic conformal motion

$$L_{\xi} g_{\mu\nu} = \xi_{\mu;\nu} + \xi_{\nu;\mu} = 2\phi g_{\mu\nu}, \tag{1.4}$$

where L denotes Lie differentiation, $g_{\mu\nu}$ is the metric tensor of the space-time, and $\phi = \frac{1}{4}\xi^{\mu}{}_{;\mu}$ is a scalar constant. We designate a vector field corresponding to $\phi = 0$, a trivial homothetic vector field (a Killing vector field). In analogy to the cases involving a Killing vector field, we define a homothetic bivector (H·B·V) according to the relation

$$\omega_{\mu;\nu} = \tfrac{1}{2}(\xi_{\mu;\nu} - \xi_{\nu;\mu}), \tag{1.5}$$

which we assume to be non-null and is generated by a homothetic vector field which is also assumed to be non-null.

The problem that we wish to investigate is to seek under various assumptions, some general properties of source-free Einstein–Maxwell equations, with a non-null electromagnetic field, admitting a non-null homothetic vector field, Eq. (1.4) generating a homothetic bivector, Eq. (1.5), which is also non-null. The analysis made here is to some extent similar to the cases involving a vacuum space-time, admitting a Killing vector field generating a non-null Killing bivector.[4,5]

II. BASIC FORMALISM

In this section we briefly summarize some of the relevant well-known tetrad formalism often used in expressing arbitrary geometric objects.[15–17] As a basis for our four-dimensional space-time we choose a tetrad of null vectors $\{\mathbf{e}_a, a = 1,2,3,4\}$ in which $\mathbf{e}_1$ and $\mathbf{e}_2$ are real and $\mathbf{e}_3$ and $\mathbf{e}_4$ are complex. The set of complex null tetrad induces a metric

 0022-2488/90/020401-09$03.00

$$g_{ab} = g^{ab} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{bmatrix}, \tag{2.1}$$

which is obtained from the inner product of basis vectors $\mathbf{e}_a \cdot \mathbf{e}_b = g_{ab}$. From the basis vectors $\mathbf{e}_a$, and its dual set θ^a, one can form a six-dimensional vector space appropriate for expressing bivectors B_{ab}. Thus if B_{ab}, is an arbitrary bivector, it may be expressed in terms of the basis θ^A, as follows:

$$B = B_A \theta^A, \quad A = \text{I,II,...,VI}, \tag{2.2}$$

where θ^A are defined as the exterior product of the dual basis θ^a. We have

$$\theta^{\text{I}} = 2\theta^4 \wedge \theta^2, \quad \theta^{\text{IV}} = 2\theta^3 \wedge \theta^2 = \overline{\theta^{\text{I}}},$$
$$\theta^{\text{II}} = 2(\theta^3 \wedge \theta^4 + \theta^2 \wedge \theta^1),$$
$$\theta^{\text{V}} = 2(\theta^4 \wedge \theta^3 + \theta^2 \wedge \theta^1) = \overline{\theta^{\text{II}}},$$
$$\theta^{\text{III}} = 2\theta^1 \wedge \theta^3, \quad \theta^{\text{VI}} = 2\theta^1 \wedge \theta^4 = \overline{\theta^{\text{III}}}. \tag{2.3}$$

The complex conjugate of a real geometrical object is obtained by performing the permutation 1234→1243, on the tetrad indices. The choice of θ^A is subject to the duality constraint $\theta^A(a\mathbf{e}_B) = \delta^A_B, A,B, = \text{I,II,...,VI}$, where the dual basis $\mathbf{e}_A$ are defined by the relationships:

$$\mathbf{e}_{\text{I}} = \mathbf{e}_4 \wedge \mathbf{e}_2, \quad \mathbf{e}_{\text{IV}} = \mathbf{e}_3 \wedge \mathbf{e}_2 = \bar{\mathbf{e}}_{\text{I}},$$
$$\mathbf{e}_{\text{II}} = \tfrac{1}{2}(\mathbf{e}_3 \wedge \mathbf{e}_4 + \mathbf{e}_2 \wedge \mathbf{e}_1),$$
$$\mathbf{e}_{\text{V}} = \tfrac{1}{2}(\mathbf{e}_4 \wedge \mathbf{e}_3 + \mathbf{e}_2 \wedge \mathbf{e}_1) = \bar{\mathbf{e}}_{\text{II}},$$
$$\mathbf{e}_{\text{III}} = \mathbf{e}_1 \wedge \mathbf{e}_3, \quad \mathbf{e}_{\text{VI}} = \mathbf{e}_1 \wedge \mathbf{e}_4 = \bar{\mathbf{e}}_{\text{III}}. \tag{2.4}$$

Accordingly, a bivector B^A, may be considered a six-vector with respect to the basis defined by Eqs. (2.3) and (2.4). One can similarly construct a metric $g_{AB} = \mathbf{e}_A \cdot \mathbf{e}_B$, which can be used to raise or lower indices on an arbitrary bivector. We have in fact

$$g_{AB} = \mathbf{e}_A \cdot \mathbf{e}_B = \begin{bmatrix} \Lambda & 0 \\ 0 & \Lambda \end{bmatrix}, \tag{2.5}$$

where

$$\Lambda = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -\frac{1}{2} & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad \Lambda^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \tag{2.6}$$

Thus if B is an arbitrary real bivector, it may be written in the form

$$B = B_A \theta^A = B_{ab} \theta^a \wedge \theta^b, \tag{2.7}$$

where the quantities $B_{\text{I}}, B_{\text{II}}, \ldots, B_{\text{VI}}$, are defined as follows:

$$B_{\text{I}} = B_{42}, \qquad B_{\text{IV}} = \overline{B}_{\text{I}} = B_{32},$$
$$B_{\text{II}} = \tfrac{1}{2}(B_{21} - B_{43}), \quad B_{\text{V}} = \overline{B}_{\text{II}} = \tfrac{1}{2}(B_{21} - B_{34}), \tag{2.8}$$
$$B_{\text{III}} = B_{13}, \qquad B_{\text{VI}} = \overline{B}_{\text{III}} = B_{14}.$$

From the bivector B, and its dual $*B$, one may form the bivectors $B^{\pm} = B \mp i{*B}$, which are self-dual and anti-self-dual, respectively. Thus the expression for the anti-self-dual B^-, has the form

$$B^- = B_A^- \theta^A = 2B_{\text{I}} \theta^{\text{I}} + 2B_{\text{II}} \theta^{\text{II}} + 2B_{\text{III}} \theta^{\text{III}}. \tag{2.9}$$

It is not difficult to show that the expression

$$2I = 4(B_{\text{I}} B_{\text{III}} - B_{\text{II}}^2), \tag{2.10}$$

obtained from evaluating $g^{AB} B_A^- B_B$, is invariant under the group of the tetrad transformations

$$\mathbf{e}_1' = \mathbf{e}^{-A} |1 - \alpha\beta|^{-1} (\mathbf{e}_1 + \beta\bar{\beta}\mathbf{e}_2 - \bar{\beta}\mathbf{e}_3 - \beta\mathbf{e}_4),$$
$$\mathbf{e}_2' = \mathbf{e}^{A} |1 - \alpha\beta|^{-1} (\alpha\bar{\alpha}\mathbf{e}_1 + \mathbf{e}_2 - \alpha\mathbf{e}_3 - \bar{\alpha}\mathbf{e}_4),$$
$$\mathbf{e}_3' = \mathbf{e}^{-iB} |1 - \alpha\beta|^{-1} (-\bar{\alpha}\mathbf{e}_1 - \beta\mathbf{e}_2 + \mathbf{e}_3 + \bar{\alpha}\beta\mathbf{e}_4),$$
$$\mathbf{e}_4' = \mathbf{e}^{iB} |1 - \alpha\beta|^{-1} (-\alpha\mathbf{e}_1 - \bar{\beta}\mathbf{e}_2 + \alpha\bar{\beta}\mathbf{e}_3 + \mathbf{e}_4), \tag{2.11}$$

where A, B, α, and β are parameters with $\alpha\beta \neq 1 \cdot A$, and B, are real, while α, and β are complex.

According to Eq. (2.10), the two independent invariants are (a): $I = 0$, and (b): $I \neq 0$. The case $I = 0$, which is known as a null bivector may be characterized by $\mathrm{B_I} = \mathrm{B_{III}} = 0$. Hence, a null bivector may be represented by its anti-self-dual bivector as

$$B^- = 2B_{\text{III}} \theta^{\text{III}} = 4B_{13} \theta^1 \wedge \theta^3. \tag{2.12}$$

Similarly, if B is a non-null bivector, it may be represented in canonical form by its corresponding anti-self-dual bivector

$$B^- = 2B_{\text{II}} \theta^{\text{II}} = 4B_{\text{II}} (\theta^3 \wedge \theta^4 + \theta^2 \wedge \theta^1), \tag{2.13}$$

where B_{II}, is given by Eq. (2.8). As it has been pointed out earlier, we will be working exclusively with the bivectors that are non-null by assumption.

For a later development, we also need the covariant derivative of a vector, denoted by a semicolon, with respect to a basis vector $\mathbf{e}_b$, as is defined by the relation

$$V_{a;b} = V_{a,b} - \Gamma_{acb} V^c, \tag{2.14}$$

where Γ_{abc} are the Ricci rotation coefficients with the properties $\Gamma_{(ab)c} = 0$, and $\Gamma_{abc} = g_{ad} \Gamma^d{}_{bc}$. Similarly,

$$V^a{}_{;b} = V^a{}_{,b} + \Gamma^a{}_{bc} V^c. \tag{2.15}$$

The Ricci tensor and the Ricci scalar R are, respectively,

$$R_{ab} = R^c{}_{abc}, \quad R = R^a{}_a, \tag{2.16}$$

while the Weyl conformal curvature tensor, excluding the cosmological term, may be expressed as

$$C_{abcd} = R_{abcd} + g_{a[c} R_{d]b} + g_{b[d} R_{c]a} - (R/6)(g_{ac} g_{bd} - g_{ad} g_{bc}). \tag{2.17}$$

Finally, if f is a scalar function, the commutation relations

$$[\mathbf{e}_a, \mathbf{e}_b] f = f_{;ab} - f_{;ba} = f_{;c} (\Gamma^c{}_{ab} - \Gamma^c{}_{ba}), \tag{2.18}$$

must hold as an integrability condition on f.

III. HOMOTHETIC VECTOR FIELD

A space-time is said to admit a homothetic vector field, if there exists an infinitesimal generator ζ^a, such that

$$\zeta_{a;b} + \zeta_{b;a} = 2\phi g_{ab}, \tag{3.1}$$

where $\phi = \frac{1}{4}\zeta^a{}_{;a}$ is a scalar constant. If $\phi = 0$, then ζ^a, is called a trivial homothetic or a Killing vector field. Equation

(3.1), when combined with the defining equation (1.5), for a homothetic bivector, may be written in the form

$$\zeta_{a;b} = \phi g_{ab} + \omega_{ab}. \tag{3.2}$$

It may be shown that the compatibility conditions for Eq. (3.1) assumes the same form as those satisfied by a Killing vector field, namely,

$$\zeta_{a;b;c} = \zeta^{d} R_{dcab}. \tag{3.3}$$

It follows from the symmetries of the Riemann tensor that

$$\zeta_{a;}{}^{b}{}_{;b} = R_{ab}\zeta^{b}, \quad \text{where } R_{ab} = R^{c}{}_{abc}. \tag{3.4}$$

Equations (3.2)–(3.4) give the results

$$\omega^{ab}{}_{;b} = R^{ab}\zeta_b = j^a, \tag{3.5}$$

with the property

$$j^a{}_{;a} = 0. \tag{3.6}$$

The vector j^a may be interpreted as a current vector generated by the (H·B·V). We also note from Eq. (3.5), that in vacuum or any space-time with $R_{ab}\zeta^b = 0$, the homothetic vector satisfies Maxwell-like equations.

The components of the Ricci tensor may be expressed in terms of the components of the electromagnetic energy-momentum tensor T_{ab}. With a suitable choice of units, T_{ab}, may be written in the form

$$T_{ab} = 2\begin{bmatrix} -F_{\rm III}\overline{F}_{\rm III} & -F_{\rm II}\overline{F}_{\rm II} & \overline{F}_{\rm II}F_{\rm III} & F_{\rm II}\overline{F}_{\rm III} \\ -F_{\rm II}\overline{F}_{\rm II} & -F_{\rm I}\overline{F}_{\rm I} & \overline{F}_{\rm I}F_{\rm II} & F_{\rm I}\overline{F}_{\rm II} \\ \overline{F}_{\rm II}F_{\rm III} & \overline{F}_{\rm I}F_{\rm II} & -\overline{F}_{\rm I}F_{\rm III} & -\overline{F}_{\rm II}F_{\rm II} \\ F_{\rm II}\overline{F}_{\rm III} & F_{\rm I}\overline{F}_{\rm II} & -F_{\rm II}\overline{F}_{\rm II} & -F_{\rm I}\overline{F}_{\rm III} \end{bmatrix}, \tag{3.7}$$

where

$$F_{\rm I} = F_{42}, \quad \overline{F}_{\rm I} = F_{32},$$
$$F_{\rm II} = \tfrac{1}{2}(F_{21} - F_{43}), \quad \overline{F}_{\rm II} = \tfrac{1}{2}(F_{21} - F_{34}), \tag{3.8}$$
$$F_{\rm III} = F_{13}, \quad \overline{F}_{\rm III} = F_{14}.$$

Lemma 3.1: Suppose a non-null homothetic vector field ζ^a, admitted by the combined Einstein–Maxwell system, generates a (H·B·V), which is also non-null, then ζ^a, is not a geodesic tangent.

Proof: Let ζ^a, be a geodesic tangent, then along the trajectory we must have

$$\zeta_{a;b}\zeta^b = \alpha\zeta_a, \tag{3.9}$$

for some scalar α. From Eqs. (3.2) and (3.9) we obtain

$$\omega_{ab}\zeta^b = (\alpha - \phi)\zeta_a. \tag{3.10}$$

Multiplying Eq. (3.10) by ζ^a, and remembering that ζ^a, is non-null by assumption, we obtain the conditions for ζ^a, to be a geodesic tangent as

$$\alpha = \phi, \tag{3.11}$$

$$\omega_{ab}\zeta^b = 0. \tag{3.12}$$

At this point we may use the transformation freedom, Eq. (2.11), to reduce the homothetic bivector ω_{ab}, to its canonical form, where $\omega_{\rm I} = \omega_{\rm III} = 0$, and $\omega_{\rm II} = \tfrac{1}{2}(\omega_{21} - \omega_{43}) \neq 0$. This can be achieved by choosing the two real null directions $\mathbf{e}_1$ and $\mathbf{e}_2$ as the principle null directions of the bivector ω_{ab}. The two bivectors ω_{ab}, and F_{ab} do not necessarily have principle null directions in common.

From Eq. (3.12), we have as a necessary and sufficient condition for a nontrivial solution ζ^a, to exist, the relationship

$$\det \omega_{ab} = 0. \tag{3.13}$$

Evaluating the above expression we obtain

$$2\omega_{21}^2\omega_{43}^2 = 0. \tag{3.14}$$

Since ω_{ab}, is by assumption non-null, the quantities ω_{21} and ω_{43} are not simultaneously zero. We therefore have the following two separate cases to consider. We have either

$$\omega_{43} = 0, \quad \omega_{21} \neq 0 \tag{3.15}$$

or

$$\omega_{21} = 0, \quad \omega_{43} \neq 0. \tag{3.16}$$

The consideration of these cases are quite similar, and therefore, it will suffice to investigate either one of them. Considering Eq. (3.15), along with Eq. (3.12) we obtain

$$\zeta_1 = \zeta_2 = 0. \tag{3.17}$$

In addition, Eqs. (1.5), (2.14), (3.1), (3.2), and (3.17) yield:

$$\omega_{12} = -\phi + \Gamma_{412}\zeta_3 + \Gamma_{312}\zeta_4, \tag{3.18}$$
$$\omega_{12} = \phi - \Gamma_{421}\zeta_3 - \Gamma_{321}\zeta_4, \tag{3.19}$$
$$\zeta_{3,1} = \Gamma_{341}\zeta_3, \tag{3.20}$$
$$\zeta_{4,1} = \Gamma_{431}\zeta_4, \tag{3.21}$$
$$\zeta_{3,4} = -\phi + \Gamma_{344}\zeta_3, \tag{3.22}$$
$$\zeta_{4,3} = -\phi + \Gamma_{433}\zeta_4, \tag{3.23}$$
$$\zeta_{3,3} = \Gamma_{343}\zeta_3, \tag{3.24}$$
$$\zeta_{4,4} = \Gamma_{434}\zeta_4. \tag{3.25}$$

We now evaluate the expressions $(\zeta_3\zeta_4)$, $a,a = 1,2,3,4$ by means of Eqs. (3.18)–(3.25) to obtain

$$(\zeta_3\zeta_4)_{,1} = 0, \quad (\zeta_3\zeta_4)_{,2} = 0, \quad (\zeta_3\zeta_4)_{,3} = -\phi\zeta_3,$$
$$(\zeta_3\zeta_4)_{,4} = -\phi\zeta_4. \tag{3.26}$$

The product $\zeta_3\zeta_4$ is a constant multiple of the squared length of the vector field ζ^a, which by assumption is non-null. Applying the commutation relation Eq. (2.18) to Eq. (3.26) yields

$$(\Gamma_{421} - \Gamma_{412})\zeta_3 + (\Gamma_{321} - \Gamma_{312})\zeta_4 = 0. \tag{3.27}$$

This result when combined with Eqs. (3.18) and (3.19) gives

$$2\omega_{12} = 0, \tag{3.28}$$

which is contrary to the assumption that the homothetic bivector is non-null. Therefore, the geodesic condition Eq. (3.9) cannot be fulfilled by a non-null homothetic vector field generating a non-null homothetic bivector. We also note that the results obtained thus far do not depend on any particular form of energy-momentum tensor, electromagnetic or otherwise. Analogous situations in a vacuum space-time involving either a trivial homothetic vector field (a Killing vector field), or a nontrivial homothetic vector field have been discussed by Debney[4] and McIntosh,[5] respectively. In particular, they have shown that in a vacuum space-time, admitting a Killing or a homothetic vector field and generating a non-null bivector cannot be a geodesic tangent. Our result is, therefore, a generalization of these results and may be stated as follows.

Theorem 3.1: In an arbitrary space-time vacuum or not, admitting a non-null homothetic vector field, generating a non-null homothetic bivector, the nontrivial homothetic vector field is not a geodesic tangent.

Lemma 3.2: Let F_{ab} be a source-free non-null electromagnetic field satisfying Einstein–Maxwell equations. Let ζ^a, be a non-null homothetic vector field admitted by the system for which $\mathbf{e}_1$ is a principle null direction for both bivectors F_{ab} and ω_{ab}. If the space-time is algebraically special with $\mathbf{e}_1$ as the degenerate principle null direction for the Weyl tensor, then $\mathbf{e}_1$ must be geodesic and sheer free.

Proof: With $\mathbf{e}_1$ as a principle null direction for both bivectors, a tetrad transformation Eq. (2.1), may be carried out to set $F_{\rm III} = F_{13} = 0$, and $\omega_{\rm III} = \omega_{13} = 0$. The vanishing of $F_{\rm III}$, gives by virtue of Eq. (3.7) the vanishing of the following components of the Ricci tensor, $R_{11} = R_{13} = R_{44} = 0$. With these results Eqs. (B2i) and (B2k) yield

$$2\kappa\Omega_2 = \psi_0\zeta_4 - \psi_1\zeta_1, \tag{3.29}$$

$$2\sigma\Omega_2 = \psi_0\zeta_2 - \psi_1\zeta_3, \tag{3.30}$$

where

$$\Omega_2 = \omega_{\rm II} = 1/2(\omega_{21} - \omega_{43}), \tag{3.31}$$

and $\psi_0, \psi_1,...,\psi_4$, are the Weyl's complex scalars.

From Eqs. (3.29) and (3.30), it immediately follows that if $\psi_0 = \psi_1 = 0$, we must have $\kappa = \sigma = 0$, whenever $\Omega_2 \neq 0$. In other words if the space-time is algebraically special, then the common principle null directions of the two non-null bivectors is geodesic and shear free.

The converse of Lemma 3.2 also holds true; namely if the common principle null direction of the two non-null bivectors F_{ab} and ω_{ab}, is geodesic and shear free, then the space-time must be algebraically special. The proof is a straightforward consequence of the generalization of the Goldberg–Sacks theorem[18] by Robinson and Schild.[19] The essence of this theorem may be stated as follows:

If with the tetrad $\mathbf{e}_a$, $a = 1,2,3,4,\mathbf{e}_1$, is geodesic and shear-free, and the Ricci tensor has the vanishing components $R_{11} = R_{13} = R_{14} = R_{33} = R_{44} = 0$, then $\psi_0 = \psi_1 = 0$; that is the space-time is algebraically special. In our case, the vanishing of these components of the Ricci tensor are satisfied by the assumption $\mathbf{e}_1$, being a principle null direction. Equations (3.29) and (3.30) with $\kappa = \sigma = 0$, give the result $\psi_1\zeta^2 = 0$, where ζ^2, is the square length of the non-null homothetic vector field. Therefore, with $\mathbf{e}_1$ being geodesic, shear-free and a common principle direction of F_{ab} and ω_{ab}, the space-time must be algebraically special with $\psi_0 = \psi_1 = 0$.

The results of Lemma 3.2 and its converse may be combined to give:

Theorem 3.2: Let F_{ab}, be a source-free non-null electromagnetic field satisfying Einstein–Maxwell field equations. Let ζ^a, be a homothetic vector field admitted by this system for which $\mathbf{e}_1$ is a geodesic and shear-free principle null direction for both of the non-null bivectors F_{ab} and ω_{ab}, then $\mathbf{e}_1$ is a degenerate principle null direction of the Weyl tensor.

Theorem 3.3: Let F_{ab} be a source-free non-null electromagnetic field satisfying Einstein–Maxwell field equations. Let ζ^a, be a homothetic vector field admitted by the system. If $\mathbf{e}_1$ and $\mathbf{e}_2$ are common principle null directions for both F_{ab} and ω_{ab}, then ζ^a is hypersurface orthogonal if and only if it is a trivial homothetic vector field.

Proof: The condition that the vector field ζ^a, to be hypersurface orthogonal, may be written in the form

$$\zeta \wedge d\zeta = 0. \tag{3.32}$$

Equations (3.32) with $\omega = d\zeta$ in its canonical form becomes

$$\omega_{12}\zeta_3 = 0, \tag{3.33}$$

$$\omega_{34}\zeta_1 = 0, \tag{3.34}$$

$$\omega_{34}\zeta_2 = 0. \tag{3.35}$$

Accordingly, we have the following two cases to consider. We have either

$$\omega_{34} = 0, \quad \zeta_3 = 0, \quad \zeta_4 = 0, \quad \omega_{12} \neq 0, \tag{3.36a}$$

or

$$\omega_{12} = 0, \quad \zeta_1 = 0, \quad \zeta_2 = 0, \quad \omega_{34} \neq 0. \tag{3.36b}$$

Let us consider Eqs. (3.36a) in which the homothetic bivector is real. In this case, we have from Eqs. (B2c), (B2l), and (3.36a)

$$2\mu\Omega_2 = -\psi_2\zeta_2, \tag{3.37}$$

$$2\rho\Omega_2 = -\psi_2\zeta_1, \tag{3.38}$$

$$\mu\zeta_1 - \bar{\rho}\zeta_2 = \phi, \quad \Omega_2 = \omega_{\rm II} = \tfrac{1}{2}\omega_{21}. \tag{3.39}$$

Combining Eqs. (3.37) and (3.38), we obtain

$$2\Omega_2(\mu\zeta_1 - \rho\zeta_2) = 0. \tag{3.40}$$

Since the homothetic bivector is non-null, we must have $\mu\zeta_1 - \rho\zeta_2 = 0$. The complex conjugate of this result when combined with Eq. (3.39) gives

$$\phi = 0, \tag{3.41}$$

as a condition for the vector field ζ^a to be hypersurface orthogonal. But this, of course, is the condition for ζ^a to be a trivial homothetic, i.e., a Killing vector field.

The relevant equations for the case in which $\omega_{12} = 0$ may be obtained from Eqs. (B2a), (B2g), and (3.36b). We have

$$2\pi\Omega_2 = -\psi_2\zeta_4, \tag{3.42}$$

$$2\tau\Omega_2 = -\psi_2\zeta_3, \tag{3.43}$$

$$\pi\zeta_3 + \bar{\pi}\zeta_4 = -\phi. \tag{3.44}$$

It is easy to show that in this case, where $\Omega_2 = -\tfrac{1}{2}\omega_{43}$ is pure imaginary, leads to the same result as dictated by Eq. (3.41).

Remarks: As an immediate consequence of Theorem 3.3, if the two non-null bivectors F_{ab} and ω_{ab} do not have principle null directions in common, then it may be possible for a hypersurface orthogonal vector field, which is a non-Killing vector field, to exist. The existence of such vector fields may shed additional light into the structure of space-times admitted by the combined Einstein–Maxwell theory. This situation is analogous for example, to the case in which a time-like hypersurface orthogonal Killing vector field is used to characterize the stationary or the static nature of a given space-time. It is therefore, of interest to seek space-times of the Einstein–Maxwell system admitting a non-null homothetic vector field, generating a non-null homothetic bivector, in which the two bivectors F_{ab} and ω_{ab} do not have principle null directions in common.

In Sec. IV we provide a particular solution to Einstein–Maxwell equations, admitting a non-null homothetic vector field ζ^a, generating a bivector ω_{ab} which is also non-null. But before proceeding to this particular solution, it is helpful to prove a theorem which excludes the possibility of constructing any solution to Einstein–Maxwell field equations under the assumptions of the theorem to follow.

Theorem 3.4: Let F_{ab} be a source-free non-null electromagnetic field satisfying Einstein–Maxwell field equations. Let ζ^a, be a non-null homothetic vector field generating a bivector ω_{ab}, which is also non-null. Furthermore, we assume the two bivectors F_{ab} and ω_{ab} have the geodesic $\mathbf{e}_1$ and $\mathbf{e}_2$ as their common principle null directions. Then the system of Eqs. (1.1)–(1.5) do not admit any solution, unless ζ^a, is a trivial homothetic vector field.

Proof: Using the geodesic conditions, and with the bivectors in their canonical forms, we have by means of Eqs. (B2a)–(B2l)

$$2\pi\Omega_2 = \psi_3\zeta_1 - \psi_2\zeta_4, \tag{3.45a}$$

$$0 = \psi_4\zeta_3 - \psi_3\zeta_2, \tag{3.45b}$$

$$2\mu\Omega_2 = \psi_3\zeta_3 - \psi_2\zeta_2, \tag{3.45c}$$

$$2\lambda\Omega_2 = \psi_4\zeta_1 - \psi_3\zeta_4, \tag{3.45d}$$

$$0 = \psi_0\zeta_4 - \psi_1\zeta_1, \tag{3.46a}$$

$$2\tau\Omega_2 = \psi_1\zeta_2 - \psi_2\zeta_3, \tag{3.46b}$$

$$2\sigma\Omega_2 = \psi_0\zeta_2 - \psi_1\zeta_3, \tag{3.46c}$$

$$2\rho\Omega_2 = \psi_1\zeta_4 - \psi_2\zeta_1, \tag{3.46d}$$

$$D\Omega_2 = \psi_1\zeta_4 - \psi_2\zeta_1 - \tfrac{1}{2}R_{12}\zeta_1, \tag{3.47a}$$

$$\Delta\Omega_2 = \psi_2\zeta_2 - \psi_3\zeta_3 + \tfrac{1}{2}R_{12}\zeta_2, \tag{3.47b}$$

$$\delta\Omega_2 = \psi_1\zeta_2 - \psi_2\zeta_3 + \tfrac{1}{2}R_{12}\zeta_3, \tag{3.47c}$$

$$\bar{\delta}\Omega_2 = \psi_2\zeta_4 - \psi_3\zeta_1 - \tfrac{1}{2}R_{12}\zeta_4, \tag{3.47d}$$

where $\Omega_2 = \tfrac{1}{2}(\omega_{21} - \omega_{43})$, $R_{12} = 2\phi_1\bar{\phi}_1$, and ϕ_1 is the nonzero tetrad component of a Maxwell field. Additional relationships may be obtained by combining Eqs. (3.46b)–(3.46d) to give

$$\tau\zeta_1 - \sigma\zeta_4 - \rho\zeta_3 = 0, \tag{3.48}$$

$$\pi\zeta_2 - \mu\zeta_4 - \lambda\zeta_3 = 0. \tag{3.49}$$

Applying the commutation relations Eqs. (A11a)–(A11d) to Ω_2, we find after some algebra the first three are satisfied by virtue of Eqs. (3.47a)–(3.48), and Eqs. (A6a) and (A6b). The commutator $(\bar{\delta}\delta - \delta\bar{\delta})\Omega_2$ gives us the result

$$\mu\zeta_1 - \rho\zeta_2 - \pi\zeta_3 + \tau\zeta_4 = \phi. \tag{3.50}$$

Evaluating (3.50) by means of Eqs. (3.45a)–(3.46d), yields

$$2\phi\Omega_2 = 0. \tag{3.51}$$

Since the homothetic bivector is non-null, $\Omega_2 \neq 0$, we must have $\phi = 0$, which is the condition for the vector field ζ^a to be a Killing vector field.

It may be noted that the introduction of a homothetic symmetry to Einstein–Maxwell field equations severely limits the possibility of having large classes of admissible solutions. This situation is indeed in sharp contrast with the case in which the underlying symmetry of Einstein–Maxwell theory is an element of an isometry group. In particular for a non-null, nontrivial homothetic vector field, generating a homothetic bivector which is also non-null to exist in a combined source-free Einstein–Maxwell system, then the bivectors F_{ab} and ω_{ab} cannot have principle null geodesic directions in common. As a supporting evidence for this conjecture, we provide a particular solution to our system of equations (1.1)–(1.5) in the next section.

IV. A PARTICULAR SOLUTION

In this section we construct a special source-free solution for the system of equations (1.1)–(1.5), satisfying the conditions:

(a) the null directions $\mathbf{e}_1^\mu = l^\mu$ and $\mathbf{e}_2^\mu = n^\mu$ are geodesics, hypersurface orthogonal and serve as the principle null directions of the non-null, source-free electromagnetic field, and

(b) the null tetrad of vectors $\mathbf{e}_a^\mu = (l^\mu, n^\mu, m^\mu, \bar{m}^\mu)$, $a = 1,2,3,4$, are parallelly propagated along the principle null directions l^μ and n^μ.

The particular solution to Einstein–Maxwell equations satisfying the above requirements has been found using the Cartan's equations of structure. We will not, however, present the details of the calculations here, since it has also been worked out by Tariq and Tupper[20] in Newman–Penrose formalism. (Some useful results shared by the principle null geodesics and when the tetrad of null vectors are parallelly propagated along them are also provided by Debney and Zund.[21,22]) Our problem therefore, reduces to determine whether or not Eq. (1.4) admits a nontrivial homothetic vector field, generating a homothetic bivector which is also non-null for the given metric obtained from the field equations (1.1)–(1.3).

The solutions to the Einstein–Maxwell field equations under assumptions (a) and (b) may be written in the form

$$l^\mu = \delta^\mu{}_r, \quad n^\mu = \delta^\mu{}_u, \quad m^\mu = \lambda_1\delta^\mu{}_y + \lambda_2\delta^\mu{}_z, \tag{4.1}$$

where the coordinates are x^μ: ($x^1 = u$, $x^2 = r$, $x^3 = y$, $x^4 = z$) and the functions λ_1 and λ_2 are, respectively,

$$\lambda_1 = u^n r^m/\sqrt{2}, \quad \lambda_2 = iu^m r^n/\sqrt{2}, \tag{4.2}$$

with $m = (\sqrt{3} - 1)/4$, and $n = -(\sqrt{3} + 1)/4$. The only nonvanishing spin coefficients are:

$$\rho=\bar\rho=-1/4r,\quad \mu=\bar\mu=1/4u, \tag{4.3}$$

$$\sigma=\bar\sigma=\sqrt{3}/4r,\quad \lambda=\bar\lambda=\sqrt{3}/4u, \tag{4.4}$$

and having the intrinsic derivatives:

$$D\rho=4\rho^2,\quad D\sigma=4\rho\sigma,\quad \Delta\mu=-4\mu^2,\quad \Delta\lambda=-4\lambda\mu, \tag{4.5}$$

with all other intrinsic derivatives being zero.

With these spin coefficients, suing the commutation relation equations (A11a)–(A11d) and Eqs. (A6a)–(A6j) defining the tetrad components of the homothetic vector field ζ^a ($a=1,2,3,4$) along with its compatibility conditions equations (A14)–(A18), we obtain:

$$\delta\zeta_1=\bar\delta\zeta_1=\delta\zeta_2=\bar\delta\zeta_2=0, \tag{4.6}$$

$$\Delta\zeta_1=4\mu\zeta_1, \tag{4.7}$$

$$D\zeta_2=-4\rho\zeta_2, \tag{4.8}$$

$$\Delta\zeta_3=\mu\zeta_3+\lambda\zeta_4, \tag{4.9}$$

$$\delta\zeta_3=\lambda\zeta_1-\sigma\zeta_2, \tag{4.10}$$

$$\bar\delta\zeta_3=\rho\zeta_2-\mu\zeta_1, \tag{4.11}$$

$$D\zeta_4=-\rho\zeta_4-\sigma\zeta_3. \tag{4.12}$$

The integration of these equations gives for the tetrad components of the vector field ζ^a with respect to the adopted coordinate system (u,r,y,z), the expressions:

$$\zeta_1=cu, \tag{4.13}$$

$$\zeta_2=(2\phi-c)r, \tag{4.14}$$

$$\zeta_3=d_1yu^{-n}r^{-m}-id_2zu^{-m}r^{-n}, \tag{4.15}$$

$$\zeta_4=d_1yu^{-n}r^{-m}+id_2zu^{-m}r^{-n}, \tag{4.16}$$

where c, d_1, and d_2 are constants. The constants d_1 and d_2 may be expressed in terms of the constant c and the homothetic constant ϕ,

$$d_1=(\alpha-\sqrt{2}\phi)/2,\quad d_2=\alpha/2,$$

$$\alpha=\sqrt{6}c/2+\sqrt{2}(1-\sqrt{2})\phi/2. \tag{4.17}$$

The tetrad components of the bivector ω_{ab} from Eqs. (3.2) are

$$\omega_{12}=c-\phi, \tag{4.18}$$

$$\omega_{13}=md_1yu^{-n}r^{-m-1}-ind_2zu^{-m}r^{-n-1}, \tag{4.19}$$

$$\omega_{42}=nd_1yu^{-n-1}r^{-m}-imd_2zu^{-m-1}r^{-n}, \tag{4.20}$$

$$\omega_{43}=0. \tag{4.21}$$

The complete particular solution to the system of Eqs. (1.1)–(1.4) with respect to (u,r,y,z) system of coordinates may be expressed in the following form. For the metric $g^{\mu\nu}={\bf e}_a{}^\mu{\bf e}_b{}^\nu g^{ab}$ and its inverse we have

$$g_{\mu\nu}=\begin{bmatrix}0&1&0&0\\1&0&0&0\\0&0&-u^{-2n}r^{-2m}&0\\0&0&0&-u^{-2m}r^{-2n}\end{bmatrix},$$

$$g^{\mu\nu}=\begin{bmatrix}0&1&0&0\\1&0&0&0\\0&0&-u^{2n}r^{2m}&0\\0&0&0&-u^{2m}r^{2n}\end{bmatrix}, \tag{4.22}$$

where $m=(\sqrt{3}-1)/4$, $n=-(\sqrt{3}+1)/4$. The Maxwell field, $f_{\mu\nu}=e_{a\mu}e_{b\nu}f^{ab}$, becomes

$$f_{\mu\nu}=\begin{bmatrix}0&-\frac{\cos\epsilon}{\sqrt{2}}(ur)^{-1/2}&0&0\\ \frac{\cos\epsilon}{\sqrt{2}}(ur)^{-1/2}&0&0&0\\0&0&0&\frac{\sin\epsilon}{\sqrt{2}}\\0&0&-\frac{\sin\epsilon}{\sqrt{2}}&0\end{bmatrix}, \tag{4.23}$$

where ϵ is an arbitrary real constant, and $f_{\mu\nu}$ is an electromagnetic field, in which both the electric and magnetic fields are in the radial direction. The vector field $\zeta^\mu={\bf e}^\mu_a\zeta^a$ takes the form:

$$\zeta^u=cu,\quad \zeta_u=(2\phi-c)r, \tag{4.24}$$

$$\zeta^r=(2\phi-c)r,\quad \zeta_r=cu, \tag{4.25}$$

$$\zeta^y=-c_1y,\quad \zeta_y=c_1yu^{-2m}r^{-2n}, \tag{4.26}$$

$$\zeta^z=c_2z,\quad \zeta_z=-c_2zu^{-2m}r^{-2n}, \tag{4.27}$$

where the constants c_1 and c_2 are

$$c_1=\sqrt{3}c/2-(1+\sqrt{3})\phi/2,\quad c_2=c_1+\phi. \tag{4.28}$$

The vector field ζ^μ, having the squared length

$$\zeta^2=(4\phi c-2c^2)ur-c_1^2y^2u^{-2n}r^{-2m}-c_2^2z^2u^{-2m}r^{-2n}, \tag{4.29}$$

is non-null and generates a bivector $\omega_{\mu\nu}=e_{a\mu}e_{b\nu}\omega^{ab}$, with components

$$\omega_{ur}=\phi-c, \tag{4.30}$$

$$\omega_{uy}=nc_1yu^{-2n-1}r^{-2m}, \tag{4.31}$$

$$\omega_{uz}=mc_2zu^{-2m-1}r^{-2n}, \tag{4.32}$$

$$\omega_{ry}=mc_1yu^{-2n}r^{-2m-1}, \tag{4.33}$$

$$\omega_{rz}=nc_2zu^{-2m}r^{-2n-1},\quad \omega_{yz}=0, \tag{4.34}$$

which is also non-null.

It is easy to verify that

$$\omega_{[\mu\nu;\lambda]} = 0, \tag{4.35}$$

as it should according to its defining Eq. (1.5) and the symmetries of the Riemann tensor Eq. (3.3).

In addition to the homothetic motion that we have found, the solution also admits three parameter groups of motion:

$$\eta_1^\mu = \delta_y^\mu, \quad \eta_2^\mu = \delta_z^\mu, \tag{4.36}$$

$$\eta_3^\mu = u\delta_u^\mu - r\delta_r^\mu - dy\delta_y^\mu + dz\delta_z^\mu, \tag{4.37}$$

where the constant $d = \sqrt{3}/2$ can be found from Eq. (4.28) by letting $\phi = 0$.

APPENDIX A: CONFORMAL MOTION AND ITS COMPATIBILITY CONDITIONS IN TETRAD REPRESENTATION

In the following we write the tetrad components equations of Eq. (1.4) and its integrability condition, which is generally expressed in terms of the vanishing of the Lie derivative of the conformal Weyl tensor:

$$L_\zeta C^\alpha{}_{\mu\nu\sigma} = 0. \tag{A1}$$

We choose a tetrad of null vector $l^\mu, n^\mu, m^\mu, \bar{m}^\mu$, with l^μ and n^μ real and m^μ complex. The only nonvanishing contractions are

$$l_\mu n^\mu = -m_\mu \bar{m}^\mu = 1. \tag{A2}$$

A frame defined by the inner product

$$\mathbf{e}_a{}^\mu \mathbf{e}^b{}_\mu = \delta^b{}_a, \quad \mathbf{e}^\mu{}_a \mathbf{e}^a{}_\nu = \delta^\mu{}_\nu, \tag{A3}$$

where $\mathbf{e}_a{}^\mu = (l^\mu, n^\mu, m^\mu, \bar{m}^\mu)$, induces a metric of the form

$$\eta_{ab} = \eta^{ab} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{bmatrix}, \quad a,b = 1,2,3,4. \tag{A4}$$

Equation (A1) in tetrad representation takes the form

$$\zeta_{a;b} + \zeta_{b;a} = 2\phi\eta_{ab} + \zeta^c(\gamma_{acb} + \gamma_{bca}), \tag{A5}$$

where γ_{abc} are the Ricci rotation coefficients.

Equation (A5) is the equivalent to the scalar equations

$$D\zeta_1 = (\epsilon + \bar{\epsilon})\zeta_1 - \bar{\kappa}\zeta_3 - \kappa\zeta_4, \tag{A6a}$$

$$\Delta\zeta_2 = -(\gamma + \bar{\gamma})\zeta_2 + \nu\zeta_3 + \bar{\nu}\zeta_4, \tag{A6b}$$

$$\delta\zeta_3 = \bar{\lambda}\zeta_1 - \sigma\zeta_2 - (\bar{\alpha} - \beta)\zeta_3, \tag{A6c}$$

$$\bar{\delta}\zeta_4 = \lambda\zeta_1 - \bar{\sigma}\zeta_2 - (\alpha - \bar{\beta})\zeta_4, \tag{A6d}$$

$$\Delta\zeta_1 + D\zeta_2 = 2\phi + (\gamma + \bar{\gamma})\zeta_1 - (\epsilon + \bar{\epsilon})\zeta_2 + (\pi - \bar{\tau})\zeta_3 + (\bar{\pi} - \tau)\zeta_4, \tag{A6e}$$

$$\delta\zeta_1 + D\zeta_3 = (\bar{\alpha} + \beta + \bar{\pi})\zeta_1 - \kappa\zeta_2 + (\epsilon - \bar{\epsilon} - \bar{\rho})\zeta_3 - \sigma\zeta_4, \tag{A6f}$$

$$\bar{\delta}\zeta_1 + D\zeta_4 = (\alpha + \bar{\beta} + \pi)\zeta_1 - \bar{\kappa}\zeta_2 + (\bar{\epsilon} - \epsilon - \rho)\zeta_4 - \bar{\sigma}\zeta_3, \tag{A6g}$$

$$\delta\zeta_2 + \Delta\zeta_3 = \bar{\nu}\zeta_1 - (\bar{\alpha} + \beta + \tau)\zeta_2 + (\mu + \gamma - \bar{\gamma})\zeta_3 + \bar{\lambda}\zeta_4, \tag{A6h}$$

$$\bar{\delta}\zeta_2 + \Delta\zeta_4 = \nu\zeta_1 - (\alpha + \bar{\beta} + \bar{\tau})\zeta_2 + (\bar{\mu} + \bar{\gamma} - \gamma)\zeta_4 + \lambda\zeta_3, \tag{A6i}$$

$$\bar{\delta}\zeta_3 + \delta\zeta_4 = -2\phi + (\mu + \bar{\mu})\zeta_1 - (\rho + \bar{\rho})\zeta_2 + (\alpha - \bar{\beta})\zeta_3 + (\bar{\alpha} - \beta)\zeta_4, \tag{A6j}$$

where the spin coefficients $\alpha,\beta,\gamma,\ldots$ are related to the Ricci rotations and may be expressed in the form

$$\gamma_{131} = \kappa, \quad \gamma_{132} = \tau, \quad \gamma_{133} = \sigma, \quad \gamma_{134} = \rho, \tag{A7}$$

$$\gamma_{241} = -\pi, \quad \gamma_{242} = -\nu,$$
$$\gamma_{243} = -\mu, \quad \gamma_{244} = -\lambda, \tag{A8}$$

$$\gamma_{121} = \epsilon + \bar{\epsilon}, \quad \gamma_{122} = \gamma + \bar{\gamma}, \quad \gamma_{123} = \beta + \bar{\alpha},$$
$$\gamma_{124} = \alpha + \bar{\beta}, \tag{A9}$$

$$\gamma_{121} - \gamma_{341} = 2\epsilon, \quad \gamma_{124} - \gamma_{344} = 2\alpha,$$
$$\gamma_{123} - \gamma_{343} = 2\beta,$$
$$\gamma_{122} - \gamma_{342} = 2\gamma.$$

The intrinsic derivatives are defined according to the relations

$$D\phi = \phi_{;\mu} l^\mu, \quad \Delta\phi = \phi_{;\mu} n^\mu, \quad \delta\phi = \phi_{;\mu} m^\mu, \quad \bar{\delta}\phi = \phi_{;\mu}\bar{m}^\mu, \tag{A10}$$

and are satisfied by the commutation relation

$$\Delta D - D\Delta = (\gamma + \bar{\gamma})D + (\epsilon + \bar{\epsilon})\Delta - (\tau + \bar{\pi})\bar{\delta} - (\bar{\tau} + \pi)\delta, \tag{A11a}$$

$$\delta D - D\delta = (\bar{\alpha} + \beta - \bar{\pi})D + \kappa\Delta - \sigma\bar{\delta} - (\bar{\rho} + \epsilon - \bar{\epsilon})\delta, \tag{A11b}$$

$$\delta\Delta - \Delta\delta = -\bar{\nu}D + (\tau - \bar{\alpha} - \beta)\Delta + \bar{\lambda}\bar{\delta} + (\mu - \gamma + \bar{\gamma})\delta, \tag{A11c}$$

$$\bar{\delta}\delta - \delta\bar{\delta} = (\bar{\mu} - \mu)D + (\bar{\rho} - \rho)\Delta - (\bar{\alpha} - \beta)\bar{\delta} - (\bar{\beta} - \alpha)\delta. \tag{A11d}$$

Equation (A1) can be written in the form

$$C_{abcd;p}\zeta^p + C_{pbcd}\zeta^p{}_{;a} + C_{apcd}\zeta^p{}_{;b} + C_{abpd}\zeta^p{}_{;c} + C_{abcp}\zeta^p{}_{;d}$$
$$= 2\phi C_{abcd} + \zeta^b[C_{rbcd}(\gamma_a{}^r{}_p + \gamma^r{}_{pa}) + C_{arcd}(\gamma_b{}^r{}_p + \gamma^r{}_{pb}) + C_{abrd}(\gamma_c{}^r{}_p + \gamma^r{}_{pc}) + C_{abcr}(\gamma_d{}^r{}_p + \gamma^r{}_{pd})]. \tag{A12}$$

The independent components of conformal curvature tensor may be expressed in terms of the five complex scalars $\psi_0, \psi_1, \ldots, \psi_4$. We have

$$C_{1212} = \psi_2 + \bar{\psi}_2, \quad C_{1213} = \psi_1, \quad C_{1223} = -\bar{\psi}_3,$$
$$C_{1234} = -\psi_2 + \bar{\psi}_2, \quad C_{1313} = \psi_0, \quad C_{1324} = -\psi_2,$$
$$C_{1334} = -\psi_1,$$
$$C_{2323} = \bar{\psi}_4, \quad C_{2334} = -\bar{\psi}_3,$$
$$C_{3434} = \psi_2 + \bar{\psi}_2, \quad C_{1314} = C_{1323} = C_{2324} = 0. \tag{A13}$$

Making use of Eqs. (A6a)–(A11d), and (A13), Eq. (A12) becomes

$$\zeta_1\Delta\psi_0+\zeta_2 D\psi_0-\zeta_3\bar\delta\psi_0-\zeta_4\delta\psi_0+2\psi_0(D\zeta_2-\delta\zeta_4)+2\psi_1(\delta\zeta_1-D\zeta_3)$$
$$=2\phi\psi_0+2\psi_0[(2\gamma-\mu)\zeta_1+(\epsilon-\bar\epsilon+\bar\rho)\zeta_2+(\pi-2\alpha)\zeta_3+(\bar\pi-\beta-\bar\alpha)\zeta_4]$$
$$+2\psi_1[(\beta+\bar\alpha-2\tau-\bar\pi)\zeta_1-\kappa\zeta_2+(\bar\epsilon-\epsilon+2\rho-\bar\rho)\zeta_3+\sigma\zeta_4], \tag{A14}$$
$$\zeta_1\Delta\psi_1+\zeta_2 D\psi_1-\zeta_3\bar\delta\psi_1-\zeta_4\delta\psi_1-\psi_0\Delta\zeta_4+\psi_1(D\zeta_2-\delta\zeta_4)+\psi_2(\delta\zeta_1-2D\zeta_3)$$
$$=\psi_0[(\pi+\bar\tau)\zeta_2-\lambda\zeta_3+(\gamma-\bar\gamma-\mu)\zeta_4]+\psi_1[(2\gamma-\mu)\zeta_1+(\epsilon-\bar\epsilon+\bar\rho)\zeta_2+(\pi-2\alpha)\zeta_3+\bar\pi-\bar\alpha-\beta)\zeta_4]$$
$$+\psi_2[(\bar\alpha+\beta-2\bar\pi-3\tau)\zeta_1-\kappa\zeta_2+(3\rho-\bar\rho-2\epsilon+2\bar\epsilon)\zeta_3+2\sigma\zeta_4], \tag{A15}$$
$$\zeta_1\Delta\psi_2+\zeta_2 D\psi_2-\zeta_3\bar\delta\psi_2-\zeta_4\delta\psi_2+\psi_1(\bar\delta\zeta_2-\Delta\zeta_4)+\psi_3(\delta\zeta_1-D\zeta_3)$$
$$=-2\phi\psi_2+\psi_1[\nu\zeta_1+(2\pi+\bar\tau-\alpha-\bar\beta)\zeta_2-\lambda\zeta_3+(\gamma-\bar\gamma+\bar\mu-2\mu)\zeta_4]$$
$$+\psi_3[(\bar\alpha+\beta-2\tau-\bar\pi)\zeta_1-\kappa\zeta_2+(2\rho-\bar\rho-\epsilon+\bar\epsilon)\zeta_3+\sigma\zeta_4], \tag{A16}$$
$$\zeta_1\Delta\psi_3+\zeta_2 D\psi_3-\zeta_3\bar\delta\psi_3-\zeta_4\delta\psi_3-\psi_4 D\zeta_3+\psi_3(\Delta\zeta_1-\bar\delta\zeta_3)+\psi_2(\bar\delta\zeta_2-2\Delta\zeta_4)$$
$$=\psi_2[\nu\zeta_1+(3\pi+2\bar\tau-\alpha-\bar\beta)\zeta_2-2\lambda\zeta_3+(2\gamma-2\bar\gamma+\bar\mu-3\mu)\zeta_4]$$
$$+\psi_3[(\bar\gamma-\gamma-\bar\mu)\zeta_1+(\rho-2\epsilon)\zeta_2+(\alpha+\bar\beta-\bar\tau)\zeta_3+(2\beta-\tau)\zeta_4]$$
$$+\psi_4[-(\tau+\bar\pi)\zeta_1+(\bar\epsilon-\epsilon+\rho)\zeta_3+\sigma\zeta_4], \tag{A17}$$
$$\zeta_1\Delta\psi_4+\zeta_2 D\psi_4-\zeta_3\bar\delta\psi_4-\zeta_4\delta\zeta_4+2\psi_2(\Delta\zeta_1-\bar\delta\zeta_3)+2\psi_3(\bar\delta\zeta_2-\Delta\zeta_4)$$
$$=2\phi\psi_4+2\psi_3[\nu\zeta_1+(2\pi+\bar\tau-\alpha-\bar\beta)\zeta_2-\lambda\zeta_3+(\gamma-\bar\gamma+\bar\mu-2\mu)\zeta_4]$$
$$+2\psi_4[(\bar\gamma-\gamma-\bar\mu)\zeta_1+(\rho-2\epsilon)\zeta_2+(\alpha+\bar\beta-\bar\tau)\zeta_3+(2\beta-\tau)\zeta_4]. \tag{A18}$$

APPENDIX B

In this appendix, we make use of Eqs. (1.1)–(1.4), (2.15), (2.16), (3.2)–(3.4), (3.9), and Eq. (A13), to write out Eq. (3.3) for an arbitrary source-free electromagnetic bivector and homothetic vector field. Let

$$\Omega_1\equiv\omega_{\rm I}=\omega_{42},\quad \Omega_2\equiv\omega_{\rm II}=\tfrac12(\omega_{21}-\omega_{43}),$$
$$\Omega_3\equiv\omega_{\rm III}=\omega_{13}. \tag{B1}$$

We obtain after a lengthy algebra:

$$D\Omega_1=-2\epsilon\Omega_1-2\pi\Omega_2+\psi_3\zeta_1-\psi_2\zeta_4+\tfrac12R_{24}\zeta_1-\tfrac12R_{44}\zeta_3, \tag{B2a}$$
$$\Delta\Omega_1=-2\gamma\Omega_1-2\nu\Omega_2-\psi_3\zeta_2+\psi_4\zeta_3+\tfrac12R_{22}\zeta_4-\tfrac12R_{24}\zeta_2, \tag{B2b}$$
$$\delta\Omega_1=-2\beta\Omega_1-2\mu\Omega_2-\psi_2\zeta_2+\psi_3\zeta_3-\tfrac12R_{24}\zeta_3+\tfrac12R_{22}\zeta_1, \tag{B2c}$$
$$\bar\delta\Omega_1=-2\alpha\Omega_1-2\lambda\Omega_2+\psi_4\zeta_1-\psi_3\zeta_4+\tfrac12R_{24}\zeta_4-\tfrac12R_{44}\zeta_2, \tag{B2d}$$
$$D\Omega_2=\kappa\Omega_1-\pi\Omega_3+\psi_1\zeta_4-\psi_2\zeta_1-\tfrac12R_{12}\zeta_1+\tfrac12R_{14}\zeta_3, \tag{B2e}$$
$$\Delta\Omega_2=\tau\Omega_1-\nu\Omega_3+\psi_2\zeta_2-\psi_3\zeta_3+\tfrac12R_{12}\zeta_2-\tfrac12R_{23}\zeta_4, \tag{B2f}$$
$$\delta\Omega_2=\sigma\Omega_1-\mu\Omega_3+\psi_1\zeta_2-\psi_2\zeta_3+\tfrac12R_{12}\zeta_3-\tfrac12R_{23}\zeta_1, \tag{B2g}$$
$$\bar\delta\Omega_2=\rho\Omega_1-\lambda\Omega_3-\psi_3\zeta_1+\psi_2\zeta_4-\tfrac12R_{12}\zeta_4+\tfrac12R_{14}\zeta_2, \tag{B2h}$$
$$D\Omega_3=2\epsilon\Omega_3+2\kappa\Omega_2+\psi_1\zeta_1-\psi_0\zeta_4+\tfrac12R_{13}\zeta_1-\tfrac12R_{11}\zeta_3, \tag{B2i}$$
$$\Delta\Omega_3=2\gamma\Omega_3+2\tau\Omega_2-\psi_1\zeta_2+\psi_2\zeta_3-\tfrac12R_{13}\zeta_2+\tfrac12R_{33}\zeta_4, \tag{B2j}$$
$$\delta\Omega_3=2\beta\Omega_3+2\sigma\Omega_2-\psi_0\zeta_2+\psi_1\zeta_3+\tfrac12R_{33}\zeta_1-\tfrac12R_{13}\zeta_3, \tag{B2k}$$
$$\bar\delta\Omega_3=2\alpha\Omega_3+2\rho\Omega_2+\psi_2\zeta_1-\psi_1\zeta_4-\tfrac12R_{11}\zeta_2+\tfrac12R_{13}\zeta_4. \tag{B2l}$$

Equations (B2a)–(B2l) are complete equations applicable to both homothetic and Killing bivectors in the combined Einstein–Maxwell theory admitting either a trivial or a nontrivial homothetic vector field.

A set of Maxwell-like equations can also be written down for the homothetic bivector ω_{ab}. These are:

$$D\Omega_2+\bar\delta\Omega_3=\Omega_3(2\alpha-\pi)+\kappa\Omega_1+2\rho\Omega_2-\tfrac12R_{12}\zeta_1-\tfrac12R_{11}\zeta_2+\tfrac12R_{13}\zeta_4+\tfrac12R_{14}\zeta_3, \tag{B3a}$$
$$\Delta\Omega_2+\delta\Omega_1=\Omega_1(\tau-2\beta)-\nu\Omega_3-2\mu\Omega_2+\tfrac12R_{12}\zeta_2+\tfrac12R_{22}\zeta_1-\tfrac12R_{23}\zeta_4-\tfrac12R_{24}\zeta_3, \tag{B3b}$$
$$\delta\Omega_2+\Delta\Omega_3=\Omega_3(2\gamma-\mu)+\sigma\Omega_1+2\tau\Omega_2+\tfrac12R_{12}\zeta_3-\tfrac12R_{13}\zeta_2+\tfrac12R_{33}\zeta_4-\tfrac12R_{23}\zeta_1, \tag{B3c}$$
$$\bar\delta\Omega_2+D\Omega_1=\Omega_1(\rho-2\epsilon)-\lambda\Omega_3-2\pi\Omega_2-\tfrac12R_{12}\zeta_4+\tfrac12R_{14}\zeta_2+\tfrac12R_{24}\zeta_1-\tfrac12R_{44}\zeta_3, \tag{B3d}$$

where the tetrad components of the Ricci tensor are given in terms of the components of energy-momentum tensor by Eq. (3.7).

If we now specialize to a source-free, non-null electromagnetic field in its canonical form $F_{\rm I}=F_{\rm III}=0$, $F_{\rm II}=1/2(F_{21}-F_{43})\neq0$, then the corresponding Maxwell equations are:

$$D\Phi = 2\rho\Phi, \quad \text{(B4a)}$$

$$\Delta\Phi = -2\mu\Phi, \quad \text{(B4b)}$$

$$\delta\Phi = 2\tau\Phi, \quad \text{(B4c)}$$

$$\bar{\delta}\Phi = -2\pi\Phi. \quad \text{(B4d)}$$

Using Eqs. (B4a) — (B4d) and the commutation relation Eqs. (A11a)–(A11d), we obtain a set of useful intrinsic derivative relationships of the Ricci tensor as well as the spin coefficients. We have in fact:

$$DR_{12} = 2(\rho + \bar{\rho})R_{12}, \quad \text{(B5a)}$$

$$\Delta R_{12} = -2(\mu + \bar{\mu})R_{12}, \quad \text{(B5b)}$$

$$\delta R_{12} = 2(\tau - \bar{\pi})R_{12}, \quad \text{(B5c)}$$

$$\bar{\delta} R_{12} = 2(\bar{\tau} - \pi)R_{12}, \quad \text{(B5d)}$$

with $R_{12} = 2\phi\bar{\phi}$, as the only nonvanishing tetrad component of the Ricci tensor. For the spin coefficients we have:

$$D\mu + \bar{\delta}\tau = \rho\bar{\mu} - \tau(\bar{\beta} - \alpha) + \sigma\lambda + \pi\bar{\pi} + \psi_2, \quad \text{(B6a)}$$

$$D\pi + \bar{\delta}\rho = \rho(\alpha + \bar{\beta}) - \tau\bar{\sigma}, \quad \text{(B6b)}$$

$$D\mu + \Delta\rho = \rho(\gamma + \bar{\gamma}) + \pi\bar{\pi} - \tau\bar{\tau}, \quad \text{(B6c)}$$

$$\Delta\tau + \delta\mu = \rho\bar{\nu} + \mu(\tau - \bar{\alpha} - \beta) + \pi\bar{\lambda} - \tau(\mu - \gamma + \bar{\gamma}), \quad \text{(B6d)}$$

$$\Delta\pi - \bar{\delta}\mu = -\rho\nu - \mu(\bar{\tau} - \alpha - \bar{\beta}) + \tau\lambda - \pi(\bar{\mu} - \bar{\gamma} + \gamma), \quad \text{(B6e)}$$

$$\delta\pi + \bar{\delta}\tau = \rho(\bar{\mu} - \mu) - \mu(\bar{\rho} - \rho) + \pi(\bar{\alpha} - \beta) - \tau(\bar{\beta} - \alpha), \quad \text{(B6f)}$$

$$\delta\rho = \rho(\bar{\alpha} + \beta + \tau) - \tau\bar{\rho} + \sigma\bar{\tau} + 2\pi\sigma + \psi_1, \quad \text{(B6g)}$$

$$\bar{\delta}\sigma = \sigma\bar{\tau} + 2\pi\sigma + \sigma(3\alpha - \bar{\beta}) + 2\psi_1. \quad \text{(B6h)}$$

Thus in addition to the Maxwell equation, we may use Eqs. (B5a)–(B6h) along with the $N{\cdot}P$ equations whenever we are dealing with a combined Einstein–Maxwell equations in which the electromagnetic field is source-free, non-null, and has been reduced to its canonical form.

[1] E. Nöther, Nachr. Akad. Wiss. Göttingen, II, Math. Phys. Kl. **235** (1918).
[2] W. R. Davis and M. K. Moss, Nuovo Cimento **38**, 1558 (1965).
[3] G. H. Katzin and J. Levine, J. Math. Phys. **9**, 8 (1968).
[4] G. C. Debney, J. Math. Phys. **13**, 1469 (1972).
[5] C. B. G. McIntosh, Gen. Rel. Grav. **7**, 199 (1976).
[6] G. H. Katzin, J. Levine, and W. R. Davis, J. Math. Phys. **10**, 617 (1969).
[7] C. D. Collinson and C. D. French, J. Math. Phys. **8**, 701 (1967).
[8] J. C. du Plessis, J. Math. Phys. **15**, 329 (1974).
[9] M. E. Cahill and A. H. Taub, Commun. Math. Phys. **21**, 1 (1971).
[10] A. H. Taub, *General Relativity: Papers in Honour of J. L. Synge*, edited by L. O'Raifeartaigh (Oxford U. P., London, 1972, Chap. VIII, p. 133.
[11] B. B. Godfrey, J. Gen. Rel. Grav. **3**, 3 (1972).
[12] B. K. Berger, J. Math. Phys. **17**, 1268 (1976).
[13] A. K. Raychaudnuri and S. K. Saha, Gen. Rel. Grav. **15**, 611 (1983).
[14] A. M. Faridi, J. Math. Phys. **28**, 1370 (1987).
[15] G. C. Debney and R. P. Kerr, J. Math. Phys. **10**, 1842 (1969).
[16] M. Cohen and J. Leory, J. Math. Mech. **16**, 501 (1966).
[17] G. C. Debney, J. Math. Phys. **12**, 1088 (1971).
[18] J. N. Goldberg and R. K. Sacks, Acta Phys. Polon. **22**, 13 (1962).
[19] I. Robinson and A. Schild, J. Math. Phys. **4**, 484 (1963).
[20] N. Tariq and B. O. J. Tupper, J. Gen. Rel. Grav. **6**, 345 (1975).
[21] G. C. Debney and J. Zund, Tensor, N. S. **22**, 333 (1971).
[22] G. C. Debney and J. Zund, Tensor, N. S. **25**, 53 (1972).
[23] K. Yano, *The Theory of Lie Derivatives and its Applications* (North-Holland, Amsterdam, 1957).
[24] E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

Electromagnetic and gravitational perturbation of type D space-times

F. Fayos
Grup de Relativitat, Societat Catalana de Fisica (I.E.C.), Barcelona, Spain and Departament de Fisica Aplicada, Universitat Politéctnica de Catalunya, Barcelona, Spain
J. J. Ferrando
Grup de Relativitat, Societat Catalana de Fisica (I.E.C.), Barcelona, Spain and Departament de Fisica Teorica, Facultat de Fisiques, Burjassot, Valéncia, Spain
X. Jaén
Grup de Relativitat, Societat Catalana de Fisica (I.E.C.), Barcelona, Spain and Departament de Fisica Fonamental, Universitat de Barcelona, Barcelona, Spain

(Received 10 May 1989; accepted for publication 6 September 1989)

A procedure to solve gauge and tetrad invariant conservation equations coupling electromagnetic and gravitational perturbations of Kerr and Kerr–Newman space-times is presented. This procedure has two steps. In the first place it must be proved that one can use the gauge and tetrad freedom to write the conservation equation as a source-free Maxwell-like equation, and second, the conservation equation has to be solved in this particular gauge.

I. INTRODUCTION

In order to study small electromagnetic and gravitational perturbations of a black hole (BH) we must find the general solution of the Maxwell–Einstein equations in the linearized theory.

If we are successful in this task, problems like the stability of a BH under small perturbations, the emission of radiation from a BH due to some small falling body, or other astrophysical problems could be analyzed.

(a) In the case of a Kerr BH, Teukolsky[1] and Teukolsky and Press,[2] using the Newman–Penrose (NP) formalism, gave two plus two equations for the Maxwell NP scalars ϕ_{0B} and ϕ_{2B} [we call these equations Teukolsky–Press (TP) relations]. The ones appearing in Ref. 1 are two decoupled separable equations in the Boyer–Linquist coordinates.

Following these works, Chandrasekhar[3] proposed an approach to determine the general solution of the Maxwell equation and he also was able to find the solution for the gravitational perturbation problem.

Recently a geometric interpretation of TP relations has been given,[4] concluding that for an arbitrary space-time, they are not a complete set of equations. On the other hand, it was also shown that once (ϕ_{0B},ϕ_{2B}) have been determined as a solution of the completed TP relations, a procedure to find ϕ_{1B} can be given. In other words, a method to find the general solution of the Maxwell equations was supplied.

(b) In the case of a Kerr–Newman BH, up to now, nobody has been able to find the general solution of the Maxwell–Einstein equations in linearized theory. The main difficulty is due to the coupling of the electromagnetic and gravitational perturbations.

Several approaches have been proposed, but the results have always been incomplete. Chandrasekhar,[5] Lee,[6] and Crossman,[7] using the NP formalism, were able to find decoupled and separable equations for some perturbed quantities.

Using a different approach, Crossman[8] has used the Cahen–Debever–Defrise complex vectorial formalism,[9] finding three decoupled equations for the three components of a self-dual complex two-form C, if C satisfies a conservation equation, i.e., $dC = 0$. Using the previous approach and the gauge and tetrad freedom, Crossman found three decoupled equations for ϕ_{0B}, ϕ_{1B}, and ϕ_{2B} from source-free Maxwell equations (recall that in a charged case the Maxwell equations have the form of a conservation equation like $dC = 0$, but with C not self-dual).

Using Crossman's result and the gauge and tetrad freedom, Crossman and Fackerell[10] and Fackerell[11] found three decoupled equations for ψ_{1B}, ψ_{2B}, and ψ_{3B} (NP scalars of the Weyl tensor) from a conservation equation, not self-dual, derived from the Maxwell–Einstein equations.

The work of Fackerell[11] is also a good review of all perturbations of Schwarzschild, Reissner–Nordstrom, and Kerr black holes.

(c) The conservation equations worked by Crossman and Fackerell have the generic form $dC_B = 0$, with $C_B = \chi_{IB} Z^I_A + \Lambda_B$ in a Kerr or a Kerr–Newman space-time. These conservation equations are directly deduced from the Maxwell–Einstein equations, that is to say, involving only algebraic manipulation but not differentiation. On the contrary, the decoupled equations for χ_{IB} were derived from these conservation equations by differentiation. This means that, in principle, the derived decoupled equations are only necessary but not sufficient conditions for the conservation equations.

The present work starts from this problem and shows a procedure to find the general solution of each one of the conservation equations in Ref. 11 for the quantities one wants to solve. We will prove that in all cases there exist a gauge and tetrad choice in such a way that the conservation equations can be written in the form $d(\chi'_{IB} Z^I_A) = 0$. After this we will solve these equations partially using the results of Ref. 4.

The paper has the following structure:

In Sec. II we introduce the complex vectorial formalism. In Sec. III we give the expressions of the NP scalars in the

 0022-2488/90/020410-06$03.00

linearized theory. We also give the transformation rules of these quantities under general gauge and tetrad transformations. In Sec. IV we show our procedure for solving a generic conservation equation. In Sec. V we apply this procedure to each one of the conservation equations. Finally, in Sec. VI, we compare our procedure with the one of Fackerell and Crossman.

II. COMPLEX VECTORIAL FORMALISM

We consider a null tetrad $\{l^\mu, n^\mu, m^\mu, \bar{m}^\mu\}$ satisfying the usual conditions

$$l_\mu n^\mu = -m_\mu \bar{m}^\mu = 1, \quad l_\mu m^\mu = n_\mu m^\mu = 0, \tag{2.1}$$

where greek indices run from 0 to 3.

We write the basis one-forms

$$l = l_\mu\, dx^\mu, \quad n = n_\mu\, dx^\mu, \quad m = m_\mu\, dx^\mu, \quad \bar{m} = \bar{m}_\mu\, dx^\mu, \tag{2.2}$$

and the self-dual basis two-forms

$$Z^0 = \bar{m} \wedge n, \quad Z^1 = n \wedge l - \bar{m} \wedge m, \quad Z^2 = l \wedge m, \tag{2.3}$$

where the self-dual character means

$$*Z^I = jZ^I, \tag{2.4}$$

where $j^2 = -1$, $*$ is the Hodge operator with $*(l \wedge n \wedge m \wedge \bar{m}) = j$, and capital indices I, J, K, run from 0 to 2.

We define the following one-forms in terms of NP quantities:

$$\sigma_0 = \tau l + \kappa n - \rho m - \sigma \bar{m}, \tag{2.5a}$$

$$\sigma_1 = \gamma l + \varepsilon n - \alpha m - \beta \bar{m}, \tag{2.5b}$$

$$\sigma_2 = \nu l + \pi n - \lambda m - \mu \bar{m}. \tag{2.5c}$$

The first equations of structure are

$$dZ^0 = -2\sigma_1 \wedge Z^0 - \sigma_2 \wedge Z^1, \tag{2.6a}$$

$$dZ^1 = 2\sigma_0 \wedge Z^0 - 2\sigma_2 \wedge Z^2, \tag{2.6b}$$

$$dZ^2 = 2\sigma_1 \wedge Z^2 + \sigma_0 \wedge Z^1. \tag{2.6c}$$

Equation (2.6b) can be written in the alternative form

$$dZ^1 = -h \wedge Z^1, \tag{2.7}$$

where the one-form h is self-defined. In terms of NP quantities,

$$h = 2(-\mu l + \rho n + \pi m - \tau \bar{m}). \tag{2.8}$$

The second equations of structure are

$$\Sigma_0 = d\sigma_0 - 2\sigma_1 \wedge \sigma_0, \tag{2.9a}$$

$$\Sigma_1 = d\sigma_1 + \sigma_0 \wedge \sigma_2, \tag{2.9b}$$

$$\Sigma_2 = d\sigma_2 + 2\sigma_1 \wedge \sigma_2, \tag{2.9c}$$

where the Σ_I are the complex curvature two-forms, which may be expanded as

$$\Sigma_I = C_{IJ} Z^J + \tfrac{1}{6} R \gamma_{IJ} Z^J + E_{IJ} \bar{Z}^J, \tag{2.10}$$

where R is the Ricci scalar, E_{IJ} is the Ricci tensor projected on the null tetrad, and

$$\gamma_{IJ} = \begin{pmatrix} 0 & 0 & \frac{1}{2} \\ 0 & -\frac{1}{4} & 0 \\ \frac{1}{2} & 0 & 0 \end{pmatrix}, \tag{2.11}$$

$$C_{IJ} = \begin{pmatrix} \psi_0 & \psi_1 & \psi_2 \\ \psi_1 & \psi_2 & \psi_3 \\ \psi_2 & \psi_3 & \psi_4 \end{pmatrix}, \tag{2.12}$$

are, respectively, the induced metric on the space of self-dual forms and the projected Weyl tensor on the null tetrad.

The Bianchi identities can be written

$$d\Sigma_0 = 2\sigma_1 \wedge \Sigma_0 - 2\sigma_0 \wedge \Sigma_1, \tag{2.13a}$$

$$d\Sigma_1 = \sigma_2 \wedge \Sigma_0 - \sigma_0 \wedge \Sigma_2, \tag{2.13b}$$

$$d\Sigma_2 = -2\sigma_1 \wedge \Sigma_2 + 2\sigma_2 \wedge \Sigma_1. \tag{2.13c}$$

The source-free Maxwell equations are, in this formalism,

$$dF = 0, \tag{2.14}$$

where $F = \phi_I Z^I$, and the ϕ_I are the Maxwell scalar fields.

Finally, the Einstein equations for the electrovacuum case are

$$R = 0, \quad E_{IJ} \bar{Z}^J = \phi_I \bar{F}. \tag{2.15}$$

These are the relevant equations we are going to use in the following. We will deal with them on a charged Petrov type D space-time. This kind of space-time can be characterized choosing a null tetrad satisfying

$$\kappa = \sigma = \lambda = \nu = 0, \tag{2.16a}$$

$$\psi_0 = \psi_1 = \psi_3 = \psi_4 = 0, \tag{2.16b}$$

$$\phi_0 = \phi_2 = 0. \tag{2.16c}$$

III. PERTURBATIONS OF SPACE-TIME

We will deal with first-order perturbations of a given space-time. We will use the following notation. No subindex A or B under a quantity means that this quantity is defined on the perturbed space-time. An A subindex under a quantity means that this quantity is defined on the unperturbed space-time, i.e., on the background space-time. A B subindex under a quantity means that this quantity is a first-order one.

The null tetrad basis one-form on the perturbed space-time $\omega^i = \{l, n, m, \bar{m}\}$ can be separated as

$$\omega^i = \omega^i_A + \omega^i_B, \tag{3.1}$$

and we can expand ω^i_B taking as basis one-forms on the perturbed space-time precisely ω^i_A. Then we will have

$$\omega^i_B = b^i_j \omega^j_A, \tag{3.2}$$

where b^1_1, b^1_2, b^2_1, and b^2_2 are real and $b^1_3 = \bar{b}^1_4$, $b^2_3 = \bar{b}^2_4$, $b^3_1 = \bar{b}^4_1$, $b^3_2 = \bar{b}^4_2$, $b^3_3 = b^4_3$, and $b^3_4 = \bar{b}^4_4$ are complex functions. For the self-dual basis two-forms we have

$$Z^I = Z^I_A + Z^I_B, \tag{3.3}$$

where now

$$Z^I_B = B^I_J Z^J_A + \underline{B}^I_J \bar{Z}^J_A \tag{3.4}$$

and the coefficients B are simply related to the coefficients b.

All the equations we can write on the perturbed space-time can be broken in two pieces. One of them is equal to the corresponding equation on the background or A space-time and the other is the first order of the equation on the A space-time.

We assume that the null tetrad on the A space-time is fixed at the beginning. We have only the freedom to perform order-B tetrad rotation and coordinate translations. We then add the B terms of such transformations in the original B terms. In other words, only the B quantities change under an order-B rotation and translation, leaving order-A quantities unchanged.

The rules for these order-B transformations are the following.

Under a general order-B rotation of the tetrad, with parameters A,v real and a,b complex arbitrary order-B functions, we obtain

$$\begin{aligned} l'^{\mu}_B &= l^{\mu}_B - A l^{\mu}_A + \bar{b} m^{\mu}_A + b \bar{m}^{\mu}_A, \\ n'^{\mu}_B &= n^{\mu}_B + A n^{\mu}_A + \bar{a} m^{\mu}_A + a \bar{m}^{\mu}_A, \\ m'^{\mu}_B &= m^{\mu}_B + a l^{\mu}_A + b n^{\mu}_A + j v \bar{m}^{\mu}_A; \end{aligned} \tag{3.5}$$

$$\begin{aligned} Z'^{0}_{B} &= Z^{0}_{B} + (A - jv) Z^{0}_{A} - \bar{a} Z^{1}_{A}, \\ Z'^{1}_{B} &= Z^{1}_{B} - 2b Z^{0}_{A} - 2\bar{a} Z^{2}_{A}, \\ Z'^{2}_{B} &= Z^{2}_{B} - b Z^{1}_{A} - (A - jv) Z^{2}_{A}; \end{aligned} \tag{3.6}$$

$$\begin{aligned} &\psi'_{0B} = \psi_{0B}, \quad \psi'_{2B} = \psi_{2B}, \quad \psi'_{4B} = \psi_{4B}, \\ &\psi'_{1B} = \psi_{1B} + 3b\psi_{2A}, \quad \psi'_{3B} = \psi_{3B} + 3a\psi_{2A}; \end{aligned} \tag{3.7}$$

$$\begin{aligned} &\phi'_{1B} = \phi_{1B}, \\ &\phi'_{0B} = \phi_{0B} + 2b\phi_{1A}, \quad \phi'_{2B} = \phi_{2B} + 2\bar{a}\phi_{1A}. \end{aligned} \tag{3.8}$$

Under a general order-B translation,

$$x'^{\mu} = x^{\mu} + \xi^{\mu}, \tag{3.9}$$

with $\xi^{\mu} = \chi l^{\mu} + Y n^{\mu} + Z m^{\mu} + \overline{Z m}^{\mu}$, where X and Y are real functions and Z is a complex arbitrary order-B function, we obtain

$$Q'_B = Q_B - L(\xi) Q_A, \tag{3.10}$$

for any quantity Q. Here $L(\xi)$ is the Lie derivative with respect to the field ξ.

We will call both, tetrad and coordinate transformations, gauge transformations.

IV. GAUGE INVARIANT PERTURBED CONSERVATION EQUATIONS

In order to determine gravitational and/or electromagnetic perturbed quantities in a Kerr or Kerr–Newman A space-time, Fackerell[11] showed that it is possible to find gauge invariant conservation equations from order-B Maxwell–Einstein equations. One can see that in all cases these equations take the generic form

$$dC_B = 0, \tag{4.1}$$

where $C_B = \chi_{IB} Z^{I}_{A} + \Lambda_B$, with χ_{IB} the perturbed quantities we want to find and Λ_B a complex general two-form not involving χ_{IB}.

In the following we present a procedure to obtain the general solution of (4.1) for a particular choice of gauge.

(a) Since the conservation equations are gauge invariant, we can write in some particular gauge (′)

$$dC'_B = d(\chi'_{IB} Z^{I}_{A} + \Lambda'_B) = 0. \tag{4.2}$$

We now expand Λ'_B in terms of Λ_B and the gauge parameters a, b, A, v, and ξ defined in (3.5) and (3.9). Because of the invariant character of (4.1), this is the same as expanding χ_{IB} in (4.1) in terms of χ'_{IB} and the gauge parameters. We can then write (4.2) in the form

$$d(\chi'_{IB} Z^{I}_{A} - G_{IB} Z^{I}_{A} + \Lambda_B) = 0, \tag{4.3}$$

where the G_{IB} are defined in terms of the properties of transformation of χ_{IB}, that is $\chi'_{IB} = \chi_{IB} + G_{IB}$, and, in general, depend linearly on the gauge parameters but not on their derivatives.

If we could choose a particular gauge in which G_{IB} would satisfy the equation

$$d(G_{IB} Z^{I}_{A}) = d\Lambda_B, \tag{4.4}$$

then (4.3) would simply turn out to be $d(\chi'_{IB} Z^{I}_{A}) = 0$.

Since (4.4) has always a solution (for every Λ_B) in the space of complex functions, a sufficient condition to make sure that this choice of gauge is possible is the following: the relation between G_{IB} and the parameters, considered as a system of linear equations for the latter, has to be compatible independently of the values of G_{IB}.

When this choice of gauge is possible, Eq. (4.1) becomes (now removing the primes from χ)

$$d(\chi_{IB} Z^{I}_{A}) = 0. \tag{4.5}$$

From now on we will call Eq. (4.5) a source-free Maxwell-like equation.

(b) In order to integrate Eq. (4.5), a method that has proved to be useful in some cases consists of deducing from (4.5) equations for χ_{0B} and χ_{2B} partially decoupled and separable in radial and angular parts.[1,2,7,8] But, assuming that these equations hold, do they guarantee the existence of a solution for Eq. (4.5)?

To solve this important question we follow a different path based on Ref. 4.

A system $D(\chi_{0B},\chi_{2B}) = 0$ is said to be a conditional system for (4.5) if all their solutions can be completed with χ_{1B} being $(\chi_{0B},\chi_{1B},\chi_{2B})$ a solution of (4.5), and all solutions of (4.5) also verify $D(\chi_{0B},\chi_{2B}) = 0$.

It was shown in Ref. 4 that a system like (4.5) admits a conditional system of second order in (χ_{0B},χ_{2B}) if and only if $dh_{\Lambda} = 0$.

In this case the conditional system is

$$\Omega(\chi_{0B},\chi_{2B}) = 0, \tag{4.6}$$

where Ω is a complex two-form with a component identically vanishing; that is to say, (4.6) consists of five equations for χ_{0B} and χ_{2B}. Their explicit expression can be found in the Appendix (in the present case with $J_B = 0$).

For each solution (χ_{0B},χ_{2B}) of (4.6) there exists a family of solutions χ_{1B} that completes the solution of (4.5). If χ_{1Bp} is a particular solution of (4.5) determined from a particular solution of (4.6) for (χ_{0B},χ_{2B}), then the family of solutions χ_{1B} corresponding to that particular solution (χ_{0B},χ_{2B}) is $\chi_{1B} = \chi_{1Bp} + \chi^{0}_{1B}$, where χ^{0}_{1B} is the general solution of the equation $d \ln \chi^{0}_{1B} = h_A$.

This general procedure is useful with the only condition being that the A space-time fulfills $dh_A = 0$.

(c) If the A space-time is a vacuum type D space-time, we can choose a null tetrad following the null directions of the Weyl tensor. Then $dh_A = 0$ holds. In this case the five equations of the conditional system have the following structure: (i) two decoupling equations for χ_{0B}, χ_{2B} [see (A8) and (A10) in the Appendix.] that coincide with the Teukolsky equations[1]; (ii) two more equations [see (A9) and (A11) in the Appendix], mixing χ_{0B}, χ_{2B}, that coincide with the ones appearing in the work of Teukolsky and Press[2]; and (iii) another equation, mixing χ_{0B}, χ_{2B}, that was found in Ref. 4 for first time.

(d) In a Kerr A space-time and working with the Boyer–Linquist coordinates, the two first TP relations can be separated into angular and radial parts by taking

$$\chi_{I'B} = R(r)\Theta(\theta)e^{j(\sigma t + m\varphi)}, \quad I' = 0,2. \tag{4.7}$$

Then we can find χ_{0B} and χ_{2B} but a constant of relative normalization. Chandrasekhar[3] shows that, given a solution for $\chi_{0B}(\chi_{2B})$, $\chi_{2B}(\chi_{0B})$ and the constant of relative normalization can be determined from one of the two second TP relations.

For fields of type (4.7), the fifth equation of the conditional system holds provided that the others hold. Then the Chandrasekhar procedure is useful for solving the general solution of the conditional system for fields like (4.7) and in a Kerr A space-time.

For each solution (χ_{0B}, χ_{2B}) we can find the family χ_{1B} that completes the general solution of (4.5) using the above procedure.

It is easy, but tedious, to prove that the whole procedure is also suitable for the same kind of equation [(5.1)] in a Kerr–Newman A space-time.

We have described the whole procedure to solve source-free Maxwell-like equations in a Kerr or a Kerr–Newman A space-time.

(e) Finally we want to remark that, if the fields we find do not factorize in the way specified in a Kerr or a Kerr–Newman A space-time or if we simply consider another A space-time (always with $dh_A = 0$), we must take into account the whole conditional system, i.e., it is not sufficient to deal only with the four TP relations; we must add the fifth equation.

V. APPLICATIONS

In this section we apply the general procedure shown in the preceding section in order to solve each one of the conservation equations presented in Ref. 11.

The steps we must follow are (i) identification of the quantities χ_{IB} and Λ_B in the conservation equation we want to deal with; and (ii) determination of the functions G_{IB} as functions of the gauge parameters analyzing their mutual independence [it is necessary that the sufficient condition quoted in Sec. IV, paragraph (a) hold].

If we are successful in these two steps we could write the conservation equation as (4.5). Then we can easily apply the general procedure shown in paragraphs (b), (c), (d), or (e) of the preceding section to find the general solution.

(a) We start with the simplest conservation equation, i.e., the source-free Maxwell equation on a vacuum A space-time

$$d[\phi_{IB} Z^I_A] = 0. \tag{5.1}$$

This system is the equivalent of Eq. (4.5). We can now follow the same procedure to solve it.

(b) The second case we will deal with is the source-free Maxwell equation on a charged type D A space-time:

$$d[\phi_{IB} Z^I_A + \phi_{1A} Z^1_B] = 0. \tag{5.2}$$

In this case we have $\chi_{IB} = \phi_{IB}$ and $\Lambda_b = \phi_{1A} Z^1_B$. Since this equation is gauge invariant, we can write it in a particular gauge ($'$):

$$d[\phi'_{IB} Z^I_A + \phi_{1A} Z'^1_B] = 0. \tag{5.3}$$

We can expand Z'^1_B in terms of Z^1_B, Z^I_A, and the gauge functions. This is completely equivalent to expanding ϕ_{1B} in (5.2) in terms of ϕ'_{IB}, ϕ_{1A}, and the gauge functions, i.e.,

$$\phi_{IB} = \phi'_{IB} - G_{IB}, \tag{5.4}$$

where

$$\begin{aligned} G_{0B} &= 2b\phi_{1A}, \\ G_{1B} &= -L(\xi)\phi_{1A} \\ &= -i(\xi)d\phi_{1A} = -i(\xi)h_A\phi_{1A}, \\ G_{2B} &= 2\bar{a}\phi_{1A}. \end{aligned} \tag{5.5}$$

We have used that $d\phi_{1A} = h_A\phi_{1A}$ holds in any charged type D A space-time.

Then we always can find a gauge ($'$) where Eq. (5.3) takes the form

$$d[\phi'_{IB} Z^I_A] = 0, \tag{5.6}$$

i.e., an equation of the general type (4.5).

(c) The third case we will present is the conservation equation for $\psi_{(I+1)B}$ on a vacuum type D A space-time[11]:

$$d[\psi_{2A}^{-1/3}\psi_{(I+1)B} Z^I_A + \tfrac{3}{2}\psi_{2A}^{2/3} Z^1_B] = 0. \tag{5.7}$$

Then we have $\chi_{IB} = \psi_{2A}^{-1/3}\psi_{(I+1)B}$ and $\Lambda_B = \frac{3}{2}\psi_{2A}^{2/3} Z^1_A$. This is again a gauge invariant equation.

We now have

$$\begin{aligned} G_{0B} &= 3b\psi_{2A}^{2/3}, \\ G_{1B} &= -\psi_{2A}^{-1/3}L(\xi)\psi_{2A} \\ &= -\psi_{2A}^{-1/3}i(\xi)d\psi_{2A} = -\psi_{2A}^{2/3}i(\xi)h_A, \\ G_{2B} &= 3a\psi_{2A}^{2/3}, \end{aligned} \tag{5.8}$$

where we have used that $d\psi_{2A} = \frac{3}{2}h_A\psi_{2A}$ holds in any vacuum type D A space-time.

Then we always can choose a gauge ($'$) where Eq. (5.7) takes the form

$$d[\psi_{2A}^{-1/3}\psi_{(I+1)B} Z^I_A] = 0, \tag{5.9}$$

that is, an equation of the general type (4.5).

(d) Finally we will deal with the conservation equation for $\psi_{(I+1)B}$ in a Kerr–Newman A space-time[11]

$$\begin{aligned} d[&\phi_{1A}^{-1/2}\psi_{(I+1)B} Z^I_A + \tfrac{3}{2}\phi_{1A}^{-1/2}\psi_{2A} Z^1_B \\ &+ 2\phi_{1A}^{-1/2}\overline{F}_B - \bar{\phi}_{1A}^{-1/2}F_B] = 0. \end{aligned} \tag{5.10}$$

Then $\chi_{IB} = \phi_{1A}^{-1/2}\psi_{(I+1)B}$ and

$$\Lambda_B = \tfrac{3}{2}\phi_{1A}^{-1/2}\psi_{2A} Z^{1B} + 2\phi_{1A}^{-1/2}\overline{F}_B - \bar{\phi}_{1A}^{-1/2}F_B.$$

Following the same steps as in (b) and (c), the G_{IB} are now

$$G_{0B} = 3b\phi_{1A}^{-1/2}\psi_{2A},$$
$$G_{1B} = -\phi_{1A}^{-1/2}L(\xi)\psi_{2A} = -\phi_{1A}^{-1/2}i(\xi)d\psi_{2A}$$
$$= -\tfrac{3}{2}\phi_{1A}^{-1/2}\psi_{2A}i(\xi)h_A - \phi_{1A}^{1/2}\bar{\phi}_{1A}^{-1/2}i(\xi)h^0_A, \qquad (5.11)$$
$$G_{2B} = 3\bar{a}\phi_{1A}^{-1/2}\psi_{2A}.$$

Then we can choose a gauge (′) where (5.10) takes the simpler form

$$d\,[\phi_{1A}^{-1/2}\psi'_{(I+1)B}Z^I_A] = 0. \qquad (5.12)$$

We have shown that in an appropriate gauge (different for each case) we can write each one of the conservation equations of Ref. 11 as a source-free Maxwell-like equation. Then we can proceed to solve these equations with the aid of the method explained in Sec. IV.

Finally we note that the conservation equation derived by Belleza and Ferrari[12] cannot be treated with our procedure, since, in this case, $G_{IB} = 0$, i.e., the required quantities are gauge invariant.

VI. CONCLUDING REMARKS

In order to derive decoupled equations for gravitational and electromagnetic perturbations of BH, Crossman and Fackerell gave a procedure applied to conservation equations.

In this section we will analyze the analogies and differences between the Crossman and Fackerell procedure and ours.

In both cases one analyzed the same kind of equation: the perturbed conservation equation

$$dC_B = 0, \quad C_B = \chi_{IB}Z^I_A + \Lambda_B. \qquad (6.1)$$

The two-form C_B will not be self-dual, in general.

For any gauge election

$$C'_B = \chi_{IB}Z^I_A - G_{IB}Z^I_A + \Lambda_B, \qquad (6.2)$$

(6.1) gives

$$d\,[\chi'_{IB}Z^I_A] = J_B, \qquad (6.3)$$

where

$$J_B \equiv d\,[-\Lambda_B + G_{IB}Z^I_A]. \qquad (6.4)$$

The Crossman and Fackerell procedure consists of operating with $*_A(d - h_A\wedge)*_A$ on both members of (6.3) and then taking only the self-dual part, i.e.,

$$(*_A(d-h_A\wedge)*_A d\,[\chi_{IB}Z^I_A], Z^J_A)$$
$$= (*_A(d-h_A\wedge)*_A J_B, Z^J_A). \qquad (6.5)$$

They proved that there exists a gauge in which the right-hand side of this equation vanishes, i.e., a gauge fulfilling

$$(*_A(d-h_A\wedge)*_A J_B, Z^J_A) = 0. \qquad (6.6)$$

The left-hand side consists of three decoupled equations (in any type D A space-time) for χ'_{IB} corresponding to three values of J. Two of these equations coincide with those found by Teukolsky. The last one, for χ'_{1B}, is the Fackerell–Ipser equation.[13]

In order to find the general solution of (6.1) the Crossman and Fackerell procedure has two problems: (a) relative normalization between χ_{0B} and χ_{2B} remains unknown; and (b) we cannot know which solution of the Fackerell–Ipser equation corresponds to a particular (χ_{0B},χ_{2B}) solution.

In our procedure, the gauge is computed from $J_B = 0$. Equation (6.1) becomes simply

$$d\,[\chi'_{IB}Z^I_A] = 0, \qquad (6.7)$$

and we have found the conditional system by applying the operator

$$\Theta B \equiv j(d - h_A\wedge)i(*_A B)Z^1_A, \qquad (6.8)$$

i.e., the conditional system is

$$\Theta(d\,[\chi'_{IB}Z^I_A]) = 0. \qquad (6.9)$$

It is easy to prove that, in type D A space-time, the Z^0_A and Z^2_A components of (6.9) coincide with the Crossman and Fackerell equations for χ'_{0B} and χ'_{2B}.

We have proved that the conditional system provides a general solution for χ'_{0B} and χ'_{2B} to the conservation equation. Then our procedure is more efficient than that of Crossman and Fackerell. What is the relation between both elections of gauges?

Taking into account our gauge equation, $J_B = 0$, we easily check that our gauge is a particular family of the Crossman and Fackerell gauge. This more accurate restriction of the gauge functions allows us to compute the general solution for the original equation.

Finally we remark that our procedure, as that of Crossman and Fackerell does not solve simultaneously (that is, in the same gauge) gravitational and electromagnetic perturbations in Kerr–Newman A space-time. This is a question that seems far from being answered and would demand new efforts.

ACKNOWLEDGMENTS

This work has been supported by Comision Asesora de Investigacion Cientifica y Tecnica under Contract No. 0046-87.

APPENDIX: THE CONDITIONAL SYSTEM

In this Appendix we give the non-null components of the conditional system for Maxwell-like equations,

$$d(\chi_{IB}Z^I_A) = J_B, \qquad (A1)$$

where J_B is a complex closed three-form.

The useful generalization of directional derivatives are

$$\Delta^{rs}_{pq} = \Delta - (p-1)\gamma + (q+1)\mu - (r-1)\bar{\gamma} + s\bar{\mu}, \qquad (A2)$$

$$D^{rs}_{pq} = D + (p+1)\epsilon - (q+1)\rho + (r-1)\bar{\epsilon} - s\bar{\rho}, \qquad (A3)$$

$$\bar{\delta}^{rs}_{pq} = \bar{\delta} - (p+1)\alpha + (q+1)\pi + (r-1)\bar{\beta} - s\bar{\tau}, \qquad (A4)$$

$$\delta^{rs}_{pq} = \delta + (p-1)\beta - (q+1)\tau - (r-1)\bar{\alpha} + s\bar{\pi}. \qquad (A5)$$

Fackerell[11] used the operators N_{cs},

$$N_{cs}f = [D^{2\;1}_{2(1-s),1-(s+c)}\Delta^{1\;0}_{1+2s,s-c}$$
$$- \delta^{2\;1}_{2(s+1),1-(s+c)}\delta^{1\;0}_{1+2s,s-c}$$
$$- (s-1)(2s-1)\psi_2]f. \qquad (A6)$$

In a vacuum type D space-time, we have

$$N_{cs}f=\psi_2^{(c-s)/3}N_{ss}[\psi_2^{-(c-s)/3}f], \tag{A7}$$

and, in a charged type D space-time,

$$N_{cs}f=\phi_1^{(c-s)/2}N_{ss}[\phi_1^{-(c-s)/2}f]. \tag{A8}$$

In both cases N_{ss} becomes a separable operator.

Let $\{\Omega_I,\underline{\Omega}_I\}$ be the components of the conditional system two-form Ω. The non-null components of the conditional system for χ_{0B} and χ_{2B} are, in any A space-time with $dh_A=0$,

$$-\Omega_0\equiv D_{0A}\chi_{0B}+D_{2A}\chi_{2B}+J_{0B}=0, \tag{A9}$$

$$-\underline{\Omega}_0\equiv \underline{D}_{0A}\chi_{0B}+\underline{D}_{2A}\chi_{2B}+J_{2B}=0 \tag{A10}$$

$$\Omega_2\equiv\widetilde{D}_{0A}\chi_{0B}+\widetilde{D}_{2A}\chi_{2B}-\widetilde{J}_{0B}=0, \tag{A11}$$

$$\underline{\Omega}_2\equiv\widetilde{\underline{D}}_{0A}\chi_{0B}+\widetilde{\underline{D}}_{2A}\chi_{2B}-\widetilde{J}_{2B}=0, \tag{A12}$$

$$\tfrac{1}{2}(\Omega_1+\underline{\Omega}_1)\equiv D_{1A}\chi_{2B}-\widetilde{D}_{1A}\chi_{0B}-J_{1B}=0, \tag{A13}$$

where

$$\begin{aligned}
D_0&=-N_{11}-\varkappa\nu+\sigma\lambda,\\
D_2&=-2\varkappa\delta^{3/2}_{3/2}{}^{1/2}_{1/2}+2\sigma D^{3/2}_{3/2}{}^{1/2}_{1/2}-\delta\varkappa+D\sigma,\\
\underline{D}_0&=\bar{\delta}^0_2{}^0_2\bar{\delta}^1_3{}^0_0-\lambda D^0_2{}^0_2\bar{\sigma}\Delta^1_3{}^0_0-\bar{\varkappa}\nu-D\lambda,\\
\underline{D}_0&=-D^0_2{}^0_2D^1_3{}^0_0-\varkappa\bar{\delta}^0_2{}^0_2-\bar{\varkappa}\delta^1_3{}^0_0-\sigma\bar{\sigma}-\bar{\delta}\varkappa,\\
D_1&=D^2_2{}^0_1\delta^1_3{}^0_0-(\tau+\bar{\pi})D^1_3{}^0_0\\
&\quad+\varkappa\Delta^2_2{}^0_1+\sigma(\pi+\bar{\tau})+\Delta\varkappa,
\end{aligned} \tag{A14}$$

and $\tilde{}$ is the operator that permutes separately the real and complex vectors of the null tetrad.

Finally J_{IB} is

$$\begin{aligned}
J_0&=\varkappa j_1+\delta^2_0{}^1_1j_2+\sigma j_3+D^2_0{}^1_1j_4,\\
J_2&=\bar{\varkappa}j_1+\bar{\delta}^0_2{}^0_2j_2-D^0_2{}^0_2j_3-\bar{\sigma}j_4,\\
J_1&=D^2_2{}^0_1j_1+\Delta^2_2{}^0_1j_2+(\bar{\pi}+\tau)j_3-(\pi+\bar{\tau})j_4,
\end{aligned} \tag{A15}$$

in terms of $j_B=j_{iB}\omega^i_B$, where j_B is the source term in the Maxwell-like equation (we have omitted the B subindex for J and j and the A subindex for all other quantities).

[1]S. A. Teukolsky, J. Appl. Phys. **185**, 635 (1973).
[2]S. A. Teukolsky and W. H. Press, J. Appl. Phys. **193**, 443 (1974).
[3]S. Chandrassekhar, *The Mathematical Theory of Black Holes* (Oxford U.P., London, 1983).
[4]B. Coll, F. Fayos, and J. Ferrando, J. Math. Phys. **28**, 1075 (1987).
[5]S. Chandrassekhar, Proc. R. Soc. London Ser. A **358**, 421 (1978).
[6]C. H. Lee, Research Institute for Fundamental Physics, Kyoto University preprint PIFP-427, 1981.
[7]R. G. Crossman, Lett. Math. Phys. **1**, 105 (1976).
[8]R. G. Crossman, Ph.D. dissertation, Monash University, June 1978.
[9]M. Cahen, R. Debever, and L. Defrise, J. Math. Mech. **16**, 761 (1967).
[10]R. G. Crossman and E. D. Fackerell, *Proceedings of the Einstein Centenary Summer School on Gravitational Radiation and Collapsed Objects*, edited by C. Edwards (Springer, Berlin, 1980).
[11]E. D. Fackerell, in *Proceedings of the Second Marcel Grossman Meeting on General Relativity*, edited by R. Ruffini (North-Holland, Amsterdam, 1982).
[12]V. Belleza and V. Ferrari, J. Math. Phys. **25**, 1985 (1984).
[13]E. D. Fackerell and J. R. Ipser, Phys. Rev. D **5**, 2455 (1972).

The noisy shift map

Lee Carlson[a] and W. C. Schieve
Center for Studies in Statistical Mechanics, University of Texas at Austin, Austin, Texas 78712

(Received 28 February 1989; accepted for publication 4 October 1989)

Starting from the topological conjugacy of a chaotic discrete map with the shift map on symbol space, ideas from the Baxter eight-vertex model and a stochastic difference equation due to Falk are used to define a notion of a noisy shift map on symbol space. Its long-time behavior is determined, along with a discussion of the ramifications of this result on the effects of noise on chaotic discrete dynamical systems.

I. INTRODUCTION

The effect of noise on chaotic dynamical systems is of great interest and has been studied by many authors. The early work on this problem was carried out by Crutchfield *et al.*,[1] who studied the effect of noise on period doubling in a discrete system. Crutchfield *et al.* found that the noise introduced a gap in the bifurcation sequence, which implied at the chaotic threshold a scaling behavior in the critical exponents. Additional work was carried out by Svensmark and Samuelson[2] on the Josephson junction: They discovered that in the presence of noise and a resonant external perturbation, the bifurcation point is shifted by an amount proportional to the square of the perturbation amplitude. In addition, Wiesenfeld and McNamara[3] have studied the amplification of a small resonant periodic perturbation in the presence of noise near the period doubling threshold. Arecchi *et al.*[4] have studied the effect of noise on the forced Duffing oscillator in the region of parameter space, where different chaotic attractors coexist: They found that the noise may lead to jumps between the different basins of attraction, with the noise, induced transitions obeying simple kinetic equations. More recently and along the same lines, Kautz[5] has investigated the problem of thermally induced escape from the basin of attraction in a dc-biased Josephson junction. Kautz found that average escape time increased exponentially with inverse temperature in the low temperature limit. Last, Kapitaniak[6] has studied the behavior of the probability density function (which is obtained form the Fokker–Planck equation) of a driven nonlinear system. Kapitaniak found that in the chaotic regime corresponding to the noise-free case, multiple maxima appear in the stationary probability density function of the driven noisy system and also defined a maximal Liapunov characteristic exponent in the presence of noise. This exponent is a random number and has a corresponding probability density function. As the noise strength increases, the mean value of this exponent approaches zero. The averaged exponent as a function of the system and driving is smoother than in the noise-free case. This implies that the noise may introduce a degree of order in the chaotic system. (Similar results were found in the Belousov–Zhabotinsky reaction by Matsumoto and Tsuda.[7])

In this work we will study the effects of noise on discrete dynamical systems. This has been studied by many authors.[8] In particular, in the context of one-dimensional chaotic maps, Crutchfield and Packard[9] have studied the symbolic dynamics of chaotic maps when they are perturbed by a noise term.

The novel approach to be suggested here is to take advantage of the topological conjugacy between a chaotic discrete map and the shift map on symbol space. This topological conjugacy is described as follows: If f: $M \to M$ is a chaotic map[10] and S: $\Sigma \to \Sigma$ is the shift map, then f is topologically conjugate to S if there exists a continuous map g:$M \to \Sigma$ such that $g \circ f = s \circ g$. One then thinks of f and s as effectively the same map. Our approach here is the following: Instead of adding a noise term to the chaotic map on real space we will perturb the (conjugate) shift map by a noise term and study the effects of the noise directly on symbol space. This will be done by making use of the formalism of the eight-vertex model[11] and a stochastic difference equation due to Falk.[12] The sequences in symbol space will be viewed as infinite configurations of spins and the shift map will be viewed as the particular transfer matrix, corresponding to the weight values (1,1,0,0,1,1,0,0), which builds up the eight-vertex lattice. Using the results of Kastelyn,[13] we see that if the weights in the eight-vertex lattice are set equal to (1,1,0,0,1,1,0,0), then the transfer matrix becomes a shift operator that shifts all arrows to the left. Taking a point w_e in the neighborhood of $w_S = (1,1,0,0,1,1,0,0)$, $w_e = w_S + ew'$, and expanding the transfer matrix $T(w_e)$ in powers of e gives, to first order in e,

$$T(w_e) \approx T(w_S) + eT' = T_S + eT_S H, \tag{1}$$

where H is the Hamiltonian for the X-Y-Z model.

We can view (1) as a *perturbed shift map* $T = T_S + eT_S H$, which was obtained by assuming fluctuations in the weights, i.e., thermal fluctuations. The perturbing term (which involves H) will be viewed as a term introducing noise into the system. The presence of the nearest-neighbor Hamiltonian H in the added term motivates us to treat a simple case where the added term is taken to depend randomly on nearest neighbors and the random variable will be taken to be independent and identically distributed. (We caution the reader that what we are calling a perturbation here is not necessarily a small perturbation).

II. FALK'S DIFFERENCE EQUATION AND THE NOISY SHIFT MAP

In order to obtain concrete results and avoid the very difficult case of an infinite number of spins, we will assume

[a] Current address: Divison of Math and Science, Dakota State University, Madison, SD 57042.

 0022-2488/90/020416-05$03.00

that there is a one-dimensional lattice of spins at sites labeled by the integers $1,\ldots,N$, where N is an extremely large positive integer. Then in this model the spin at site j has the left neighbor $j-1$ and the right neighbor $j+1$. At each site a discrete time-dependent spin variable $s_j(t)$ is assigned which takes the values $(1,-1)$. In this language the strictly deterministic behavior of the original shift map is as follows: At time t the spin at site j takes on the value of the spin at site $j-1$ at time $t+1$:

$$s_j(t+1) = s_{j-1}(t). \tag{2}$$

In our model of the noisy shift map the evolution of the spin variables is given by the following stochastic process at time t: Spin $s_j(t)$ depends on the two neighbor spins $s_{j-1}(t)$ and $s_{j+1}(t)$. If the two neighbor spins are parallel, then $s_j(t+1)$ assumes the value of $s_{j-1}(t)$ [or $s_{j+1}(t)$]. If the two neighboring spins are antiparallel, then one takes a Bernoulli trial, where $s_j(t+1)$ is either spin up or spin down depending on the value of the random variable in the Bernoulli trial. This process can be described by the following nonlinear stochastic difference equation first considered by Falk in a different context[12]:

$$s_j(t+1) = \tfrac{1}{2}[s_{j-1}(t) + s_{j+1}(t)] + \tfrac{1}{2}[1 - s_{j-1}(t)s_{j+1}(t)]\,\Omega_j(t), \quad \text{for } 2 \leqslant j \leqslant N-1, \tag{3}$$

where

$$s_1(t+1) = s_2(t), \quad s_N(t+1) = s_{N-1}(t),$$

the random variables $\Omega_j(t)$ are independent and identically distributed, and

$$\text{probability}(\Omega_j(t) = +1) = \tfrac{1}{2}, \qquad \text{probability}(\Omega_j(t) = -1) = \tfrac{1}{2}. \tag{4}$$

Equations (4) mean that the coin is taken to be fair.

Define

$$s_j' = s_j(t+1), \quad s_j = s_j(t), \tag{5}$$

and let $d(s,s') = \frac{1}{2}(1+ss')$ be the Kronecker delta. The above stochastic process can be associated with a probability $P(s_1',\ldots,s_N'|s_1,\ldots,s_N)$ such that the spin system will be in a state $s_1',\ldots,s_N'$ at time $t+1$ if it was in a state $s_1,\ldots,s_N$ at time t. This probability is given by the expression[12]

$$P(s_1',\ldots,s_N'|s_1,\ldots,s_N) = \frac{1}{2}(1+s_1's_2)\Bigg\{\prod_{j=2,\ldots,N-1}\Bigg[\frac{1}{2}(1+s_{j-1}s_{j+1})\frac{1}{2}(1+s_js_{j+1})d(s_j',s_j) + \frac{1}{2}(1+s_{j-1}s_{j+1}) \times\frac{1}{2}(1-s_js_{j+1})d(s_j',-s_j) + \frac{1}{2}(1-s_{j-1}s_{j+1})\Bigg[\frac{1}{2}d(s_j',1) + \frac{1}{2}d(s_j',-1)\Bigg]\Bigg]\Bigg\}\frac{1}{2}(1+s_N's_{N-1}). \tag{6}$$

This is a one-step transition matrix and defines a Markov chain associated with the stochastic difference equation. Let $\mathbf{P}$ be the $2^N\times 2^N$ matrix with the elements $P(s_1',\ldots,s_N'|s_1,\ldots,s_N)$. Here $\mathbf{P}$ is a stochastic matrix since each of its elements are in the interval $[0,1]$ and the elements in a column sum to 1.

Since the result (heads or tails) of tossing a coin is not dependent on the value of the spin at a nearest-neighbor site, then $\langle s_i(t)\Omega_j(t)\rangle = \langle s_i(t)\rangle\langle\Omega_j(t)\rangle$ and hence from (3) it follows that[12]

$$\begin{aligned}\langle s_1(t+1)\rangle &= \langle s_2(t)\rangle,\\ \langle s_j(t+1)\rangle &= [\langle s_{j-1}(t)\rangle + \langle s_{j+1}(t)\rangle]/2\\ &\quad + [1 - \langle s_{j-1}(t)s_{j+1}(t)\rangle]\langle\Omega_j(t)\rangle,\\ \langle s_N(t+1)\rangle &= \langle s_{N-1}(t)\rangle.\end{aligned} \tag{7}$$

For a fair toss one has that $\langle\Omega_j(t)\rangle = 0$ and the system (7) becomes a closed and linear system of equations involving only single-spin averages:

$$\langle s(t+1)\rangle = A\,\langle s(t)\rangle, \tag{8}$$

where A is the $N\times N$ nonsymmetric matrix $A_{ii}=0$, $A_{12} = 1 = A_{N,N-1}$ and has all elements on either side of the diagonal equal to $\frac{1}{2}$. All other elements are equal to zero. This implies that $A-qI$ is a nonsymmetric matrix that is the same as A except that it has $A_{ii} = -q$ for every i. Define $a=-q$, $b=1$, $c=\frac{1}{2}$, and define[12] a real $N\times N$ diagonal matrix S by $S_{11} = S_{NN} = b^{1/2}$, $S_{22} = S_{33} = \cdots = S_{N-1,N-1} = c^{1/2}$, where all other elements are zero. Then S^{-1} is the matrix with $S_{11}^{-1} = S_{NN}^{-1} = b^{-1/2}$, $S_{22}^{-1} = \cdots S_{N-1,N-1}^{1} = c^{-1/2}$. From this one has that the matrix elements of the similarity transformation $S' = S^{-1}(A-qI)S$ of $A-qI$ are given by $S_{ii}' = a$ for every i, $S_{12}' = S_{21}' = S_{N-1,N}' = S_{N,N-1}' = (bc)^{1/2}$, where all other elements on either side of the diagonal are equal to c, and $S_{ij}' = 0$ for all other i,j. Therefore,[12] $S^{-1}AS$ is real and symmetric, which implies that its eigenvalues are real. By the Sturm separation theorem, $S^{-1}AS$ has N distinct eigenvalues. Hence the same theorem applies to A and therefore the eigenvalues of A are real and nondegenerate. The characteristic equation $|A-qI|_N = 0$ can be expressed as[12]

$$\sin\beta\sin(N-1)\beta = 0, \tag{9}$$

with the solutions

$$\beta_k = k\pi/(N-1), \quad (k=0,1,\ldots,N-1). \tag{10}$$

The components of the right eigenvectors $\mathbf{v}_k$ of A given by[12]

$$(\mathbf{v}_k)_j = v_{1k}\cos[(j-1)\beta_k]. \tag{11}$$

If $\mathbf{u}_k$ are the right eigenvectors of the matrix $S^{-1}AS$, then since the N eigenvalues of A are distinct we have $\mathbf{v}_k = S\mathbf{u}_k$ and $\mathbf{u}_k = S^{-1}\mathbf{v}_k$. Also, since S^{-1} is symmetric it is equal to its transpose. Therefore, taking the $\mathbf{u}_k$ gives $\mathbf{u}_k^T = \mathbf{v}_k^{T_S-1}$. The spectral representation of the matrix $S^{-1}AS$ is[12]

$$S^{-1}AS = \sum_k \frac{e_k \mathbf{u}_k \mathbf{u}_k^T}{\mathbf{u}_k^T \mathbf{u}_k}$$

$$= \sum_k \frac{e_k S^{-1} \mathbf{v}_k \mathbf{v}_k^{T_S - 1}}{\mathbf{v}_k^{T_S - 2} \mathbf{v}_k}$$

and thus the spectral representation of the matrix A is

$$A = \sum_k \frac{e_k \mathbf{v}_k \mathbf{v}_k^{T_S}}{\mathbf{v}_k^{T_S - 2} \mathbf{v}_k}. \tag{12}$$

The solution of (8) is then[12]

$$\langle \mathbf{s}(t) \rangle = \left(\sum_{k = 0,\ldots,N-1} \frac{e_k^t \mathbf{v}_k \mathbf{v}_k^{T_S - 2}}{v_k^{T_S - 2}} \mathbf{v}_k \right) \langle s(0) \rangle, \tag{13}$$

where $e_k = \cos(\beta_k)$.

Since $t \to \infty$ the only terms left in the sum are for $k = 0$ and $k = N-1$. The term with $k = 0$ corresponds to "ferromagnetic" alignment and $k = N-1$ corresponds to "antiferromagnetic" alignment. Therefore, we have[12]

$$\lim_{t\to\infty} \langle s_j(2t) \rangle$$

$$= \sum_{i=1,N} \left\{ \left(\frac{\mathbf{v}_0 \mathbf{v}_0^{T_S - 2}}{\mathbf{v}_0^{T_S - 2} \mathbf{v}_0} \right)_{ij} \langle s_i(0) \rangle + \left(\frac{\mathbf{v}_{N-1} \mathbf{v}_{N-1}^{T_S - 2}}{\mathbf{v}_{N-1}^{T_S - 2} \mathbf{v}_{N-1}} \right)_{ij} \langle s_i(0) \rangle \right\},$$

$$p^+(s^0) = \lim_{t\to\infty} p(1|s^0;t)$$

$$= \lim_{t\to\infty} [1 + \langle s_j(2t) \rangle]$$

$$= [1 + \langle s_j(\infty) \rangle]/2,$$

$$p^-(u^0) = [1 - \langle s_j(\infty) \rangle]/2, \tag{14}$$

where the limit of $\langle s_j(2t) \rangle$ as $t \to \infty$ is given by[12]

$$\langle s_j(\infty) \rangle = (N-1)^{-1} s_{N/2}^0 + 2(N-1)^{-1} \times \sum_{n = 1,\ldots,N/2-1} s_n^0. \tag{15}$$

III. THE SUPPRESSION OF CHAOS BY NOISE

In order to analyze these results and their relation to chaotic dynamics we will use the topological characterization of chaos using the formalism of symbolic dynamics.[14] In symbolic dynamics the transition matrix of a subshift of finite type tells us what sequences are allowed in symbol space. For a transition matrix given by

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \tag{16}$$

all pairs of symbols can occur: 11,01,10, and 00. However, for

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \tag{17}$$

the pair 00 cannot occur. We can represent the shift map on the set of these admissible sequences in terms of the arrow diagrams in the eight-vertex model[11] as follows. Let the transition matrix be given by (16) and let $\{x_i\}$ be an admissible sequence, say, for example,

FIG. 1. Representation of the sequence $\{x_i\}$ as an infinite configuration of spins.

$$x_i = 0, \; i < 0,$$
$$= 1, \; 0 \leqslant i \leqslant 6,$$
$$= 0, \; i > 6.$$

Define the value of $x_i = 1$ as "spin up" and $x_i = 0$ as "spin down" and fix this infinite configuration in space (see Fig. 1). Then the shift S on this sequence $\{x_i\}$ can be represented as shown in Fig. 2. As a convenient mnemonic, this diagram can also be represented in terms of eight-vertex diagrams[11] using the usual rules which follow.

(i) If two spins in the horizontal direction are parallel (antiparallel), then between them place an up (down) arrow.

(ii) If two spins in the vertical direction are pointing in the same (opposite) direction, then between them place a right-pointing (left-pointing) arrow.

The above shift operation can then be represented as shown in Fig. 3. One can continue the shift operation and its inverse to obtain an infinite two-dimensional lattice. Note that if the spin at place i has orientation x_i, then this orientation will be preserved along the diagonal of the lattice. This means that under the operation of the (deterministic) shift map the vertex diagrams shown in Fig. 4 are not allowed physically. This is not surprising since in the eight-vertex model it was learned that the shift map is the transfer matrix parametrized by the weight values (1,1,0,0,1,1,0,0), which correspond to the diagrams shown in Fig. 5. However, the noisy shift map can flip spins along the diagonal and thus introduce extra diagrams not present in the ordinary deterministic shift map. Motivated by these considerations it is instructive to make the following definitions.

Definition 1: Let $F{:}M \to M$ be a chaotic map on some set M and suppose it is topologically conjugate to the shift map on the sequences S: $\Sigma_A \to \Sigma_A$. If under a noisy perturbation of S there exist additional admissible sequences for S one says that the *chaos is enhanced by noise.*

Definition 2: If under the perturbation some of the admissible sequences do not exist after an infinite period of time, one says that the *chaos is suppressed by noise.*

The notions expressed in Definitions 1 and 2 can be formalized even more as follows. If A is an $n \times n$ transition matrix, where Σ_A is the set of admissible symbol sequences, then Σ_A is a Cantor set unless A is a permutation matrix.[15]

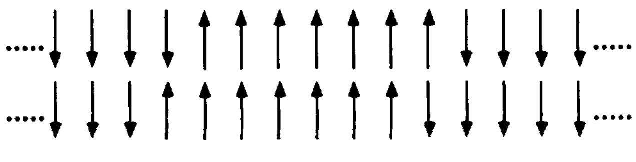


FIG. 2. Representation of the shift operation on the sequence $\{x_i\}$.

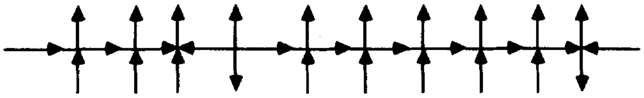


FIG. 3. Representation of the shift operation on the sequence $\{x_i\}$ in terms of eight-vertex diagrams.

FIG. 5. Allowed eight-vertex diagrams for the shift map.

This is true since a permutation matrix has only a single "1" in each row and column. If the transition matrix is a permutation matrix, then the shift on Σ_A is a single periodic orbit. One can then restate the second part of Definition 2 as follows.

Definition 3: If under the noisy perturbation the transition matrix approaches a permutation matrix after infinite time, then the chaos is suppressed by noise.

In Sec. II the Falk solution[12] of the Markov chain associated with what we interpreted as the noisy shift map was stated. It was seen that on average only the sequences in a "ferromagnetic" configuration survived. In other words, only the pairs 00 and 11 can occur and the pairs 10 or 01 cannot occur. Thus with the noise term the transition matrix approaches a permutation matrix after an infinite amount of time. Thus the chaos, as characterized *topologically*, is suppressed by the noise.

Since it is the intersecting of the stable and unstable manifolds that results in the chaotic behavior of dynamical systems, it is of interest for the present formulation to characterize these points of intersection, or *homoclinic* points, explicitly in symbol space. The stable and unstable manifolds Σ_A are characterized as follows. If $\mathbf{x}$ is a point in Σ_A, then the local stable and unstable manifolds of $\mathbf{x}$ in Σ_A are given by[10]

$$W^s_m(\mathbf{x}) = \{\mathbf{y} \text{ in } \Sigma_A | y_k = x_K \text{ for every } k \geqslant m\},$$

$$W^u_m(\mathbf{x}) = \{\mathbf{y} \text{ in } \Sigma_A | y_k = x_k \text{ for every } k \leqslant m\}.$$

The (global) stable and unstable manifolds of $\mathbf{x}$ are then given by

$$W^s(\mathbf{x}) = \lim_{m\to\infty} W^s_m(\mathbf{x}),\quad W^u(\mathbf{x}) = \lim_{m\to-\infty} W^u_m(\mathbf{x}).$$

A point $\mathbf{y}$ in Σ_A is *homoclinic* to a point $\mathbf{x}$ in Σ_A if $\mathbf{y}$ is in the intersection of $W^s(\mathbf{x})$ and $W^u(\mathbf{x})$. For example, if $\mathbf{x} = (\cdots 00 \cdots 000 \cdots)$, then if $\mathbf{y}$ is homoclinic to $\mathbf{x}$, $\mathbf{y}$ is of the form $\mathbf{y} = (\cdots 000 y_m y_{m+1} \cdots y_{n-1} y_n 000 \cdots)$ for some integers n and m. If one shifts a sequence in a homoclinic configuration using the shift map it will always remain in a homoclinic configuration. From the results in Sec. II, it is seen that a sequence in a homoclinic configuration will no longer remain so (on average) after an infinite amount of time when the shift map is perturbed by the chosen noise term. Under the action of the noisy shift map, the only configurations that remain are those that have all the symbols the same or the configuration with zero in place i and 1 in place $i+1$. Thus, again, we can say that the chaos is *suppressed by noise*.

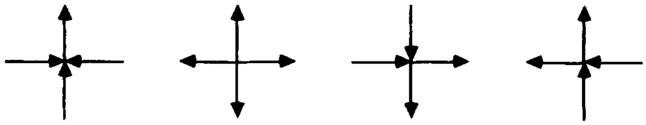


FIG. 4. Unallowed eight-vertex diagrams for the shift map.

In Ref. 16 the authors used a notion of a quantum Melnikov function to show that quantum fluctuations have the effect of suppressing the chaotic behavior in the forced Duffing oscillator. Also, in Ref. 17 Carlson used the formalism of Churchill *et al.*[18] and the notion of the Gaussian effective potential due to Stevenson[19] to show that quantum fluctuations have the effect of suppressing the chaos in the Henon–Heiles potential. Further, in Ref. 20 it is shown by using Melnikov techniques that weak Langevin noise has the effect of raising the threshold for chaotic behavior. Thus these results furnish additional examples of Definition 3.

One might wonder if there are any examples of Definition 1. In Refs. 17 and 21 it is shown that noise due to quantum fluctuations can enhance the chaotic behavior.

ACKNOWLEDGMENTS

Author WCS gratefully acknowledges conversations with Adi Bulsara.

Author LC would like to acknowledge the support of the Robert Welch Foundation.

[1]J. Crutchfield, J. D. Farmer, and B. Huberman, Phys. Rep. **92**, 45 (1982).
[2]H. Svensmark and M. Samuelson, Phys. Rev. A **36**, 2413 (1987).
[3]K. Wiesenfeld and B. McNamara, Phys. Rev. Lett. **55**, 10 (1985); Phys. Rev. A **33**, 629 (1986).
[4]F. Arrecchi, R. Badii, and A. Politi, Phys. Rev. A **32**, 402 (1985).
[5]R. Kautz, Phys. Rev. A **38**, 2066 (1988).
[6]T. Kapitaniak, *Chaos in Systems with Noise* (World Scientific, Singapore, 1988).
[7]K. Matsumoto and I. Tsuda, J. Stat. Phys. **31**, 87 (1983).
[8]See, in particular, F. Argoul, A. Arneodo, P. Collet, and A. Leane, Europhys. Lett. **3**, 643 (1987); J. P. Crutchfield, M. Nauenberg, and J. Rudnick, Phys. Rev. Lett. **46**, 933 (1981); J. Guckenheimer, Nature **298**, 358 (1982); H. Herzel, W. Ebeling, and Th. Schulmeister, Z. Naturforsch. **42**, 136 (1987); H. Herzel and B. Pompe, Phys. Lett. A **122**, 121 (1987); R. Kapral, M. Schell, and S. Fraser, J. Phys. Chem. **86**, 2205 (1982); E. Ott and J. D. Hanson, Phys. Lett. A **85**, 20 (1982); E. Ott, E. D. Yorke, and J. A. Yorke, Physica **16D**, 62 (1985); B. Shaiman, C. F. Wayne, and P. C. Martin, Phys. Rev. Lett. **46**, 935 (1981); S. Thomas and S. Grossman, J. Stat. Phys. **26**, 485 (1984); R. Vallee, C. Delisle, and C. Chrostowski, Phys. Rev. A **30**, 336 (1984); A. Zippelius and M. Lucke, J. Stat. Phys. **24**, 345 (1981).
[9]J. Crutchfield and N. Packard, Physica **7D**, 201 (1983).
[10]R. Devaney, *An Introduction to Chaotic Dynamical Systems* (Benjamin Cummings, Menlo Park, CA, 1986).
[11]R. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic, New York, 1982).
[12]This derivation has been extracted from various papers of H. Falk; see J. Math Phys. **26**, 692 (1985); Physica A **104**, 459 (1980); Physica A **117**, 561 (1983).
[13]P. W. Kasteleyn, in *Fundamental Problems in Statistical Mechanics III, Proceedings of the 1974 Wageningen Summer School*, edited by E. G. D. Cohen (North-Holland/American Elsevier, New York, 1975).
[14]J. Moser, *Stable and Random Motion in Dynamical Systems* (Princeton U. P., Princeton, NJ, 1977).

[15]R. Bowen, *On Axiom A Diffeomorphisms*, Lecture Notes in Mathematics, Vol. 470 (Springer–Verlag, New York, 1975).
[16]T. Petrosky and W. Schieve, Phys. Rev. A **31**, 3097 (1985).
[17]L. Carlson, Ph.D. thesis, University of Texas at Austin, 1989.
[18]R. C. Churchhill, G. Pecelli, and D. L. Rod, J. Diff. Eqs. **17**, 329 (1975); R. Churchhill and D. L. Rod, J. Diff. Eqs. **37**, 23 (1980).
[19]P. M. Stevenson, Phys. Rev. D **30**, 1712 (1984).
[20]A. R. Bulsara and W. C. Schieve, "Homoclinic Chaos in Systems with Noise" (submitted for publication).
[21]W. C. Schieve and T. Petrosky, in *Quantum Optics IV*, edited by J. Harvey and D. Walls (Springer, New York, 1986).

Hyperinflation in the Ising model on quasiperiodic chains

T. Odagaki[a)]
Department of Physics, Brandeis University, Waltham, Massachusetts 02254

(Received 4 October 1988; accepted for publication 13 September 1989)

Using a hyperinflation rule, the free energy of the two component Ising system on a chain with an arbitrary quasiperiodic order is shown to be given by an average of the free energy of each component, in agreement with the result obtained by the transfer matrix formalism.

I. INTRODUCTION

Quasiperiodicity has been known to give rise to various extraordinary dynamic properties that cannot be seen in either periodic or random systems.[1] On the other hand, it is believed that critical exponents of phase transitions on quasiperiodic lattices are the same as those for periodic lattices and quasiperiodic lattices belong to the universality class of periodic lattice.[2] This is a clear contrast to random systems, which may have nonuniversal critical exponents.[3]

In this paper, I use a hyperinflation rule[4] to obtain the exact expression of the free energy of the one-dimensional Ising model, where two kinds of interactions are arranged in a certain quasiperiodic order. The hyperinflation rule relates two different quasiperiodic (and periodic) sequences. Thus repeated applications of the hyperinflation transformation transform a quasiperiodic sequence to a sequence at the fixed point of the transformation. If the free energy of the sequence at the fixed point is known, one can determine the free energy of the quasiperiodic chain by inverting the transformation.

The Ising model with zero external field on a chain can be solved exactly by the transfer matrix formalism,[5] since the transfer matrices commute. Therefore, the aim of the present paper is to clarify how the hyperinflation rule can be utilized to obtain certain physical quantities such as the free energy. Although this method is very similar to the renormalization group method for regular Ising chains,[6] these two methods are essentially different in that while the renormalization group technique requires the system be self-similar after a renormalization group transformation, the present method is based on the similarity in two different sequences related by a hyperinflation.

II. THEORY

I consider the familiar Hamiltonian $H(K_0,K_1,\alpha)$ for the semi-infinite one-dimensional Ising model for spins, $\sigma_i = \pm 1$, with two kinds of nearest neighbor interactions K_0 and K_1 ($kT=1$):

$$H = -\sum_{i=1}^{\infty} K_{S(i)}\sigma_i\sigma_{i+1}, \tag{1}$$

where $S(i)$ is defined by

$$S(i) = [\alpha(i+1)] - [\alpha i], \tag{2}$$

with the Gauss symbol $[\cdots]$. The sequence of $S(i)$ consists of 0 and 1 and it is periodic when α is rational and quasiperiodic when α is irrational. It is trivial that the frequencies of 0 and 1 in the sequence are $1-\alpha$ and α, respectively. The sequence of 0 and 1 determined by Eq. (2) has hyperinflation symmetries.[4] For example, an inflation

$$0 \to 10, \quad 1 \to 1 \tag{3}$$

transforms a sequence for α to a sequence for α', which is defined by

$$\alpha' = 1/(2-\alpha). \tag{4}$$

The left fixed point of this transformation is $\alpha^* = 1$.

Now, I consider the partition function

$$Z(K_0,K_1,\alpha) = \sum_{\{\sigma_i = \pm 1\}} \exp[-H(K_0,K_1,\alpha)] \tag{5}$$

and the free energy per spin

$$f(K_0,K_1,\alpha) = -N^{-1}\ln Z(K_0,K_1,\alpha)$$

(N is the number of spins in the system and is always considered to be infinity in the thermodynamic limit) for sequence α and sequence α', which are related by the hyperinflation transformation (4). One can easily perform a partial summation for the partition function $Z(K_0,K_1,\alpha')$ and obtain the renormalization transformation

$$Z(K_0,K_1,\alpha') = e^{(1-\alpha)Ng(K_0,K_1)}Z(K_0',K_1,\alpha), \tag{6}$$

where

$$g(K_0,K_1) = \tfrac{1}{2}\ln[4\cosh(K_0+K_1)\cosh(K_0-K_1)] \tag{7}$$

and

$$\tanh K_0' = \tanh K_0 \tanh K_1. \tag{8}$$

Therefore the free energy satisfies the following recursion relation:

$$f(K_0,K_1,\alpha') = [1/(2-\alpha)][f(K_0',K_1,\alpha) - (1-\alpha)g(K_0,K_1)]. \tag{9}$$

I consider first a series of rational α's, $\alpha^{(m)} = m/(m+1)$ ($m = 0,1,\ldots$), where $\alpha^{(m+1)}$ and $\alpha^{(m)}$ are related via Eq. (4). Noting that $m=0$ corresponds to a regular chain of K_0 and hence

$$f(K_0,K_1,0) = -\ln(2\cosh K_0),$$

it is straightforward to show by induction that

$$f\left(K_0,K_1,\frac{m}{m+1}\right) = -\frac{\ln(2\cosh K_0) + m\ln(2\cosh K_1)}{m+1}. \tag{10}$$

[a)] Present address: Department of Physics, Kyoto Institute of Technology, Matsugasaki, Sakyoku, Kyoto 606, Japan.

Note that as $m \to \infty, f(K_0,K_1,m/(m+1))$ behaves as

$$f(K_0,K_1,m/(m+1)) = -\ln(2\cosh K_1) + \frac{1}{m}\ln\frac{\cosh K_0}{\cosh K_1} + O(m^{-2}). \tag{11}$$

Now, for an arbitrary α, I apply the transformation (4) n times, to obtain $\alpha,\alpha',...,\alpha^{(n)}$, for which the free energy satisfies the recursion relation

$$f(K_0^{(n-k-1)},K_1,\alpha^{(k+1)}) = [1/(2-\alpha^{(k)})][f(K_0^{(n-k)},K_1,\alpha^{(k)}) - (1-\alpha^{(k)})g(K_0^{(n-k-1)},K_1)], \tag{12}$$

where

$$\tanh K_0^{(k+1)} = \tanh K_1 \tanh K_0^{(k)}$$

or, equivalently,

$$\frac{\cosh^2 K_0^{(k+1)}}{\cosh^2 K_0^{(k)}} = \frac{\cosh^2 K_1}{\cosh(K_0^{(k)}+K_1)\cosh(K_0^{(k)}-K_1)},$$

with $K_0^{(0)} = K_0$ and $\alpha^{(0)} = \alpha$. Eliminating $f(K_0^{(n-k)}, K_1,\alpha^{(k)})$ $(k = 1,2,...,n-1)$ from this set of equations, I find

$$f(K_0^{(n)},K_1,\alpha) = (n+1-n\alpha)f(K_0,K_1,\alpha^{(n)}) + (1-\alpha)\sum_{k=0}^{n-1} g(K_0^{(k)},K_1). \tag{13}$$

Now, take the $n \to \infty$ limit. Then $\alpha^{(n)}$ converges to the fixed point, $\alpha^* = 1$, of the transformation (4) and thus $f(K_0,K_1,\alpha^{(n)})$ approaches

$$-\ln[2\cosh(K_1)] + n^{-1}\ln[(\cosh K_0)/(\cosh K_1)]$$

as in Eq. (11). Using the explicit expression of $g(K_0^{(k)},K_1)$, one can easily show that the free energy for the sequence for an arbitrary α is given by

$$f(K_0,K_1,\alpha) = -[(1-\alpha)\ln(2\cosh K_0) + \alpha\ln(2\cosh K_1)]. \tag{14}$$

This reduces consistently to Eq. (10) when $\alpha = m/(m+1)$.

III. CONCLUSION

I have obtained in this paper the exact expression of the free energy for the Ising system with binary couplings arranged in a quasiperiodic order of Eq. (2) using the hyperinflation transformation. The expression (14) indicates that the free energy is given by the simple average of the free energies of each component, in agreement with the result obtained by the transfer matrix formalism.[5] The present method can be considered as a generalized renormalization group approach, in which the transformation relates two different systems instead of self-similar systems required in the usual renormalization group method. Thus this method may be applicable to a much wider class of problems including problems in higher-dimensional quasilattices, although hyperinflation rules in two- and higher-dimensional quasilattices are yet to be found.

[1] *The Physics of Quasicrystals*, edited by P. J. Steinhardt and S. Ostlund (World Scientific, Singapore, 1987); J. Sokoloff, Phys. Rep. **126**, 189 (1985).

[2] J. P. Lu, T. Odagaki, and J. L. Birman, Phys. Rev. B **33**, 4809 (1986); J. P. Lu and J. L. Birman, J. Stat. Phys. **46**, 1057 (1987); H. Aoyama and T. Odagaki, *ibid.* **48**, 503 (1987); J. Phys. A **20**, 4985 (1987); Int. J. Mod. Phys. B **2**, 13 (1988); T. C. Choy, preprint.

[3] A. B. Harris, J. Phys. C **7**, 1671 (1974).

[4] T. Odagaki and H. Aoyama, Phys. Rev. Lett. **61**, 775 (1988).

[5] Y. Achiam, T. C. Lubensky, and E. W. Marshall, Phys. Rev. B **33**, 6460 (1986).

[6] M. Nauenberg, J. Math. Phys. **16**, 703 (1975).

Effective interactions involving massive and massless particles of spin $\frac{1}{2}$, 1, $\frac{3}{2}$, and 2

D. Spehler
Universite Louis Pasteur, IUT, 3, rue St. Paul, 67300 Strasbourg, France

(Received 6 June 1989; accepted for publication 23 August 1989)

The Bargman–Wigner theory is used in order to quantize and build interactions "à la Fermi"between massive or massless particles with spin $\frac{1}{2}$, 1, $\frac{3}{2}$, and 2.

I. INTRODUCTION

Until now only spin-$\frac{1}{2}$ and spin-1 particles have been observed as elementary particles. But we have no reason to exclude the existence of elementary spin-$\frac{3}{2}$ or spin-2 particles. On the contrary, recent unified theories such as supersymmetry (SUSY), implies the existence of such particles.

Nevertheless, SUSY has not yet won experimental support. Consequently one has to look for other approaches in order to get a unified theory; the superstring theory being one of them. Before establishing a satisfactory unified theory we must therefore be able to describe, even approximately, interactions between particles with different spins in order to give the experimentalist a way to identify these hypothetical new elementary particles. This is our motivation for proposing a general method "à la Fermi" for constructing effective interactions. We obtained a large class of interactions with this model.

To construct such interactions we first need a formalism involving both massive as well as massless particles. With the Bargmann–Wigner (BW) theory[1] as a starting point, we have recently developed its connection with the usual formalism for massive particles.[2]

In the present work we have shown how we can construct the BW theory for massless particles (Sec. II). From a two-component formalism described in Sec. III, we obtained a four-component spinor formalism, and have been able to construct interactions between particles with or without mass, by contractions of spinor indices. Before constructing the interactions, we verify in Sec. IV that our theory gives the usual results for massless particles with spins 1 and $\frac{1}{2}$. We have also constructed projection operators for spin-$\frac{3}{2}$ massless particles and calculated the sum over polarization for spin-2 massless particles. (Our notation is summarized in the Appendix.)

II. THE BW MASSLESS THEORY

Following BW's theory, we assume that a massless particle of spin s is described by a spinor

$$\psi_{a_1\cdots a_k\cdots a_{2s}}(x),\quad a_k=1,2,3,4,$$

of rank $2s$, symmetric in its spin variables, that obey a system of $2s$ Dirac's equations:

$$i(\gamma\cdot\partial)_{a_k'a_k}\psi_{a_1\cdots a_k\cdots a_{2s}}(x)=0, \tag{2.1}$$

with $k=1,...,2s$.

From Eq. (2.1) we may derive the following two equations:

$$i(\gamma\cdot\partial)_{a_1'a_1}(\psi_R)_{a_1\cdots a_{2s}}(x)=0,$$
$$i(\gamma\cdot\partial)_{a_1'a_1}(\psi_L)_{a_1\cdots a_{2s}}(x)=0, \tag{2.2}$$

where

$$\psi_R(x)=(\psi_R)_{a_1\cdots a_{2s}}(x)$$
$$=\tfrac{1}{2}(I+\gamma^5)_{a_1a_1'}\cdots\tfrac{1}{2}(I+\gamma^5)_{a_{2s}a_{2s}'}\psi_{a_s'\cdots a_{2s}'}(x),$$
$$\psi_L(x)=(\psi_L)_{a_1\cdots a_{2s}}(x)$$
$$=\tfrac{1}{2}(I-\gamma^5)_{a_1a_1'}\cdots\tfrac{1}{2}(I-\gamma^5)_{a_{2s}a_{2s}'}(x)\psi_{a_s'\cdots a_{2s}'}(x). \tag{2.3}$$

We note that symmetry in spin indices for ψ implies symmetry in spin indices for ψ_R and ψ_L.

It is clear that if ψ transforms under the Poincare group like the tensor product of $2s$ bispinors:

$$\psi\sim\begin{pmatrix}\xi_{\rm I}\\ \dot\chi^{\rm I}\end{pmatrix}\otimes\cdots\otimes\begin{pmatrix}\xi_{{\rm II}s}\\ \dot\chi^{{\rm II}s}\end{pmatrix},$$

then, in chiral representation, which we will use from now on for convenience, ψ_R and ψ_L will transform like:

$$\psi_R\sim\begin{pmatrix}\xi_{\rm I}\\ 0\end{pmatrix}\otimes\cdots\otimes\begin{pmatrix}\xi_{{\rm II}s}\\ 0\end{pmatrix},$$
$$\psi_L\sim\begin{pmatrix}0\\ \dot\chi^{\rm I}\end{pmatrix}\otimes\cdots\otimes\begin{pmatrix}0\\ \dot\chi^{{\rm II}s}\end{pmatrix}.$$

Since the spin operator in chiral representation is

$$\boldsymbol\Sigma=\tfrac{1}{2}(\boldsymbol\Sigma\otimes I\otimes\cdots\otimes I+\cdots+I\otimes\cdots\otimes\boldsymbol\Sigma)$$

(in this definition we have a sum of $2s$ terms, each of them being constituted by $2s$ tensor products) and in momentum space is

$$\boldsymbol\sigma\cdot\mathbf n\xi(p)=\xi(p),$$
$$\boldsymbol\sigma\cdot\mathbf n\dot\chi(p)=-\dot\chi(p),$$

where $\mathbf n=(\mathbf p/|\mathbf p|)$, we will have the following properties:

$$\boldsymbol\Sigma\cdot\mathbf n\psi_R(p)=s\psi_R(p),$$
$$\boldsymbol\Sigma\cdot\mathbf n\psi_L(p)=-s\psi_L(p), \tag{2.4}$$

provide that ψ transforms like the tensor product of $2s$ bispinors. Equations (2.4) means that ψ_R and ψ_L describe particles with spin $+s$ $(-s)$. Adopting the notation

$$\psi_R = \begin{pmatrix} \varphi \\ 0 \end{pmatrix}, \quad \psi_L = \begin{pmatrix} 0 \\ \dot{\omega} \end{pmatrix}, \tag{2.5}$$

where φ and $\dot{\omega}$ are the only nonzero components of ψ_R and ψ_L, and are two-component spinors of rank $2s$:

$$\varphi = \varphi_{b_1 \cdots b_{2s}}, \quad \dot{\omega} = \omega_{\dot{b}_1 \cdots \dot{b}_{2s}}, \quad b_1, b_2, \ldots, b_{2s} \dot{b}_1, \ldots, \dot{b}_{2s} = 1,2,$$

the only nontrivial equation, derived from the Eqs. (2.2) are:

$$(-i\sigma^0 \partial_0 - i\boldsymbol{\sigma} \cdot \nabla) \otimes \mathbb{1} \otimes \cdots \oplus \mathbb{1} \varphi(x) = 0, \tag{2.6}$$

$$(-i\sigma^0 \partial_0 + i\boldsymbol{\sigma} \cdot \nabla) \otimes \mathbb{1} \otimes \cdots \oplus \mathbb{1} \dot{\omega}(x) = 0, \tag{2.7}$$

$\varphi(x)$ and $\dot{\omega}(x)$ being symmetric spinors. We remark that we can derive, for example, Eq. (2.7) from Eq. (2.6) by space reflection. We can see that the only nontrivial components of a particle of spin s ($-s$) are those associated with a two-component symmetric spinsor $\varphi_{b_1 \cdots b_{2s}}(x)$ ($\omega^{\dot{b}_1 \cdots \dot{b}_{2s}}(x)$) obeying Eqs. (2.6) and (2.7), where now $b_1, \ldots, b_{2s}, \dot{b}_1, \ldots, \dot{b}_{2s} = 1,2$.

Our aim in next section is, using the properties of the two-component spinors φ and $\dot{\omega}$, to obtain a four-component formalism for ψ_R and ψ_L, that is analogous to the formalism constructed for the massive case.[2]

III. FROM TWO-COMPONENT SPINORIAL FORMALISM TO THE FOUR-COMPONENT SPINORIAL FORMALISM

The symmetry in spin indices of the two-component spinsors φ and $\dot{\omega}$ allow us to express these fields as a linear combination of the symmetric (2×2) matrices.[3] The coefficients of such expansions are the following new fields: (i) vector fields $f^k(x)$ and $f^{*k}(x)$ in the case of spin-1 particles ($k = 1,2,3$), (ii) vector-spinor fields $\eta^k_b(x)$ and $\theta^k_{\dot{b}}(x)$ in the case of spin-$\frac{3}{2}$ particles ($k = 1,2,3$, and $b = 1,2$), and (iii) tensor fields $f^{kj}(x)$ and $f^{*kj}(x)$ in the case of spin-2 particles ($k,j = 1,2,3$). We derive the properties of the new fields f^k, f^{*k}, η^k_b, $\theta^k_{\dot{b}}$, f^{kj}, and f^{*kj} from the symmetry properties of the spinor φ and $\dot{\omega}$ and from the Pauli equation (2.6) and (2.7) satisfied by φ and $\dot{\omega}$, respectively. We summarize our results in Table I. In this table, indices $b_1, \ldots, b_{2s}, \dot{b}_s, \ldots, \dot{b}_{2s}$ take the value 1,2. The variables k, j, l, take the value 1,2,3, and G_1, G_2 are constants with the dimension of mass.

We are now able to construct a four-component formalism. Recalling definition (2.3) for ψ_R and ψ_L and Eqs. (2.2) satisfied by these fields we have shown that the correspondence between two-component and four-component formalism is coherent, or in other words, that Eqs. (2.3) and apropriate decomposition for ψ_R and ψ_L give rise to the new fields f^k, f^{*k}, η^k_b, $\theta^k_{\dot{b}}$, f^{kj}, and f^{*kj}, and obeying the properties indicated in Table I.

For a spin-$\frac{1}{2}$ field we have

$$i(\gamma \cdot \partial) \psi_{\substack{R \\ L}} = 0,$$

where

$$\psi_L = \tfrac{1}{2}(I + \gamma^5)\psi = \begin{pmatrix} \varphi \\ 0 \end{pmatrix}, \tag{3.1a}$$

$$\psi_L = \tfrac{1}{2}(I - \gamma^5)\psi = \begin{pmatrix} 0 \\ \dot{\omega} \end{pmatrix}. \tag{3.1b}$$

On the other hand, for a spin-1 field we obtained

$$\{i(\gamma \cdot \partial) \otimes I\} \psi_{\substack{R \\ L}} = 0$$

whereas,

$$\psi_{\substack{R \\ L}} = \sqrt{G_1} \begin{pmatrix} f^k \\ f^{*k} \end{pmatrix} (\tfrac{1}{2}(I \pm \gamma^5)(\Sigma^k C)), \tag{3.2}$$

C being the (4×4) conjugation matrix. In the case of a spin-$\frac{3}{2}$ field we get

$$\{i(\gamma \cdot \partial) \otimes I \otimes I\} \psi_{\substack{R \\ L}} = 0,$$

$$(\psi_{\substack{R \\ L}})_{a_1 a_2 a_3} = (\Omega_{\substack{R \\ L}})^k_{a_1} (\tfrac{1}{2}(I \pm \gamma^5)(\Sigma^k C))_{a_2 a_3}, \tag{3.3a}$$

TABLE I. From BW formalism to usual fields.

Spin	Two-component spinors	Decomposition	Properties of the new fields
$\frac{1}{2}$	$\varphi_b(x)$ $\varphi^{\dot{b}}(x)$		
1	$\varphi_{b_1 b_2}(x)$	$= \sqrt{G_1} f^k(x) \cdot (\sigma^k C_1)_{b_1 b_2}$	$\nabla \cdot \mathbf{f} = 0$
	$\omega^{\dot{b}_1 \dot{b}_2}(x)$	$= \sqrt{G_1} f^{*k}(x) \cdot (\sigma^k C_1)_{b_1 b_2}$	$i\partial^0 f^k = \epsilon^{kjl} \partial_j f^l$
$\frac{3}{2}$	$\varphi_{b_1 b_2 b_3}(x)$ $\omega^{\dot{b}_1 \dot{b}_2 \dot{b}_3}(x)$	$= \eta^k_{b_1}(x) \cdot (\sigma^k C_1)_{b_2 b_3}$ $= \theta^k_{\dot{b}_1}(x) \cdot (\sigma^k C_1)_{b_2 b_3}$	$\boldsymbol{\sigma}_{b_1 \dot{b}_2} \boldsymbol{\eta}_{b_2} = \boldsymbol{\sigma}_{b_1 \dot{b}_2} \boldsymbol{\theta}_{\dot{b}_2} = 0$ $\nabla \cdot \boldsymbol{\eta}_{b_1} = \nabla \cdot \boldsymbol{\theta}_{\dot{b}_1} = 0$ $i\partial^0 \eta^k_{b_1} = \epsilon^{kjl} \partial_j \eta^k_{b_1}$ $i\partial^0 \theta^k_{\dot{b}_1} = -\epsilon^{kjl} \partial_j \theta^k_{\dot{b}_1}$ $\partial^0 \eta^k_{b_1} = -\boldsymbol{\sigma} \cdot \nabla \eta^k_{b_1}$ $\partial^0 \theta^k_{\dot{b}_1} = \boldsymbol{\sigma} \cdot \nabla \theta^k_{\dot{b}_1}$
2	$\varphi_{b_1 b_2 b_3 b_4}(x)$ $\omega^{\dot{b}_1 \dot{b}_2 \dot{b}_3 \dot{b}_4}(x)$	$= \sqrt{G_2} f^{kj}(x) \cdot (\sigma^j C_1)_{b_1 b_2} (\sigma^k C_1)_{b_3 b_4}$ $= \sqrt{G_2} f^{*kj}(x) (\sigma^j C_1)_{b_1 \dot{b}_2} \cdot (\sigma^k C_1)_{b_3 b_4}$	$f^{kk} = 0$ $f^{kj} = f^{jk}$ $\partial_k f^{kj} = 0$ $\partial_0 f^{kj} + i\epsilon^{lnj} \partial_l f^{kn} = 0$

and where we have defined $\psi^k_{\substack{R\\L}}$ as

$$(\Omega_R)^k = \tfrac{1}{2}(I+\gamma^5)\Omega^k = \begin{pmatrix}\eta^k\\0\end{pmatrix},$$
$$(\Omega_L)^k = \tfrac{1}{2}(I-\gamma^5)\Omega^k = \begin{pmatrix}0\\\theta^k\end{pmatrix}. \quad (3.3b)$$

For a spin-2 field we have

$$\{i(\gamma\cdot\partial)\otimes I\otimes I\otimes I\}\psi_{\substack{R\\L}} = 0,$$

whereas,

$$\begin{pmatrix}\psi_R\\L\end{pmatrix} = \sqrt{G_2}\begin{pmatrix}f^{kj}\\f^{*kj}\end{pmatrix}(\tfrac{1}{2}(I\pm\gamma^5)\Sigma^jC)_{a_1a_2}(\tfrac{1}{2}(I\pm\gamma^5)\Sigma^kC)_{a_3a_4}, \quad (3.4)$$

where $a_1, a_2, a_3, a_4 = 1, 2, 3, 4$.

Before constructing interactions, something that we are now technically able to do, we will first check in the next section that the formalism constructed here gives rise to the usual projectors for particles with half-integer spin and sum over the polarizations for particles with integer spin.[4] In addition we will also obtain a way to write the decompositions Eqs. (3.1)–(3.4) in a relativistic form.

IV. QUANTIZATION OF THE FIELDS

First we will quantize the nonvanishing two-component spinor fields φ (or $\dot\omega$) assuming[3] commutation relations (4.1),

$$[\varphi_{b'_1\cdots b'_{2s}}(x), \varphi^+_{b'_2\cdots b'_{2s}}(y)]_s,$$
$$= (-i)^{2s-1}K\sum_{\mathscr{P}} i(i\sigma^0\partial_0 - i\boldsymbol{\sigma}\cdot\nabla)_{b'_1b_1}\cdots$$
$$\times i(i\sigma^0\partial_0 - i\boldsymbol{\sigma}\cdot\nabla)_{b'_{2s}b_{2s}}D(x-y), \quad (4.1)$$

where $D(x-y)$ is the Jordan–Pauli function for massless particles, K is a constant to be determined, and $\mathscr{P}$ denotes all possible permutations among the spinor indices. We are using the convention (4.2)

$$[\phi,\psi]_s = \phi\psi + (-1)^{2s-1}\psi\phi, \quad (4.2)$$

s being the spin of the field. We remark that commutation rules for $\dot\omega$ follows from (4.1) by space reflection. We obtain in this way a method to derive all the properties of $\dot\omega$ fields from the properties of φ fields. When we expand the two-componet φ fields in terms of the new field f^k, η^k_b, f^{kj}; imposing (4.1) we get the commutation relations for the new fields. From now on we will treat the problem in two distinct ways, depending whether we are dealing with integer or half-integer spin. When we are interested in particles with half-integer spin we use Eqs. (3.1) and (3.3) to obtain commutation rules between the four-component fields $(\psi_R)_a$ and $(\Omega_R)^k_a$ (particles with spin-$\frac{1}{2}$ and $\frac{3}{2}$). On the other hand, we derived from the well known plane wave expansion of the field ψ_R and $(\Omega_K)^k_a$ another expression of the anticommutator for these fields. Comparing the two last expressions and adopting convenient normalizations, we were able to fix the value of the constant K and we obtained projection operators. The results for particles with four-momentum $p = (p^0, \mathbf{p})$ are summarized in Table II. In this table we used Bjorken's notations

$$\eta^\mu = (1,0,0,0),$$
$$\hat p^\mu = ((p\cdot\eta)^2 - p^2)^{1/2}(p^\mu - (p\cdot\eta)\eta^\mu).$$

However if we are interested in particles with integer spin, we will proceed in a slightly different way. For an spin-1 particle,[3,6] if we define

$$f_\mu(x) = \tfrac{1}{2}\epsilon_{0\mu\alpha\beta}\mathscr{F}^{\alpha\beta}(x) \quad (4.3)$$

(note that f_0 is necessary zero), where in (4.3)

$$\mathscr{F}^{\alpha\beta}(x) = F^{\alpha\beta}(x) - i\tilde F^{\alpha\beta}(x),$$
$$F^{\alpha\beta}(x) = \partial^\beta\mathscr{A}^\alpha(x) - \partial^\alpha\mathscr{A}^\beta(x), \quad (4.4)$$
$$\tilde F^{\alpha\beta}(x) = \tfrac{1}{2}\epsilon^{\alpha\beta\mu\nu}F_{\mu\nu}(x),$$
$$\alpha,\beta,\mu,\nu = 0,1,2,3,$$

$\mathscr{A}^\mu_{(x)}$ is the vector potential. Using the properties of the f fields given in Table I and Eqs. (4.3) and (4.4) we clearly reobtain Maxwell's equations. Postulating the usual plane wave expansion of $\mathscr{A}^\mu(x)$,

$$\mathscr{A}^\mu(x) = \frac{1}{(2\mu)^{3/2}}\int\frac{d^3p}{\sqrt{2p^0}}\sum_{\lambda=1}^{2}\epsilon^\mu(p,\lambda)$$
$$\times\{A(p,\lambda)e^{-ipx} + A^+(p,\lambda)e^{ipx}\}, \quad (4.5)$$

where

$$P = (p^0,\mathbf{p}),$$
$$[A(p,\lambda), A^+(p',\lambda')] = \delta^3(\mathbf{p}-\mathbf{p}')\delta_{\lambda\lambda'}, \quad (4.6)$$

λ and λ' are polarizations, and using (4.3) we obtain a different expression from that of the beginning of Sec. IV, for the commutator between f fields. Equalizing these two results (adopting Bjorken's[5] notation) we obtain for a particle with four-momentum p and spin 1:

$$X^{\mu\nu}(p) = \sum_{\lambda=1}^{2}\epsilon^\mu(p,\lambda)\epsilon^{\nu*}(p,\lambda)$$
$$= -g^{\mu\nu} + \eta^\mu\eta^\nu - \hat p^\mu\hat p^\nu. \quad (4.7)$$

Equation (4.7) is exactly the well known expression for sum over polarization. We derive from the relations satisfied by the f fields, given in Table I, the usual properties of the polarization vector $\epsilon^\mu(p,\lambda)$. For a spin-2 massless particle

TABLE II. Operators of projection for massless $\frac{1}{2}$ and $\frac{3}{2}$ particles.

Spin	K	Operators of projection
$\frac{1}{2}$	1	$\begin{pmatrix}P_R\\L\end{pmatrix}_{a_1a_2}(p) = \begin{pmatrix}U_R\\L\end{pmatrix}_{a_1}(p)\begin{pmatrix}U_R\\L\end{pmatrix}^+_{a_2}(p) = \frac{1}{2\lvert\mathbf{p}\rvert}\left(\frac{1}{2}(I\pm\gamma^5)p\gamma^0\right)_{a_1a_2}$
$\frac{3}{2}$	$\frac{1}{2\lvert\mathbf{p}\rvert^2}$	$\begin{pmatrix}P_R\\L\end{pmatrix}^{\mu\nu}_{a_1a_2}(p) = \begin{pmatrix}U_R\\L\end{pmatrix}_{a_1}(p)\begin{pmatrix}U_R\\L\end{pmatrix}^+_{a_2}(p) = \frac{1}{4\lvert\mathbf{p}\rvert}[\frac{1}{2}(I\pm\gamma^5)p\gamma^0]_{a_1a_2}\cdot(-g^{\mu\nu}+\eta^\mu\eta^\nu-\hat p^\mu\hat p^\nu \pm i\epsilon^{0\alpha\mu\nu}\hat p^\alpha) = \frac{1}{4\lvert\mathbf{p}\rvert}[\frac{1}{2}(I\pm\gamma^5)p\gamma^0]_{a_1a_2}Y^{\mu\nu}(p)$

we proceed in a quite analogous manner, that is, we define

$$f_{\mu\mu'}(x) = \tfrac{1}{4}\epsilon_{0\mu\alpha\beta}\epsilon_{0\mu'\alpha'\beta'}\mathscr{F}^{\alpha\beta,\alpha'\beta'}(x) \qquad (4.8)$$

(note that the only nontrivial component of f are obtained for μ and μ' are nonzero). Here

$$\mathscr{F}^{\alpha\beta,\alpha'\beta'}(x) = F^{\alpha\beta,\alpha'\beta'}(x) - i\tilde{F}^{\alpha\beta,\alpha'\beta'}(x),$$

$$\mathscr{F}^{\alpha\beta,\alpha'\beta'}(x) = \partial^{\alpha}(\partial^{\alpha'}\mathscr{A}^{\beta\beta'}(x) - \partial^{\beta'}\mathscr{A}^{\beta\alpha'}(x) - \partial^{\beta}(\partial^{\alpha'}\mathscr{A}^{\alpha\beta'}(x) - \partial^{\beta'}\mathscr{A}^{\alpha\alpha'}(x)),$$

$$\tilde{F}^{\alpha\beta,\alpha'\beta'}(x) = \tfrac{1}{4}\epsilon^{\alpha\beta\mu\nu}\epsilon^{\alpha'\beta'\mu'\nu'}F_{\mu\nu,\mu'\nu'},$$

$$\alpha,\beta,\alpha',\beta',\mu,\nu,\mu',\nu' = 0,1,2,3,$$

and $\mathscr{A}^{\mu\nu}(x)$ plays the role of the vector potential as in the case of spin-1 theory. We postulate the following plane wave expansion for $\mathscr{A}^{\mu\nu}(x)$:

$$\mathscr{A}^{\mu\nu}(x) = \frac{1}{(2\pi)^{3/2}}\int\frac{d^3p}{\sqrt{2p^0}}\sum_{\lambda=1}^{2}\epsilon^{\mu\nu}(p,\lambda) \times\{A(p,\lambda)e^{-ipx} + A^{+}(p,\lambda)e^{ipx}\}, \qquad (4.9)$$

where we used the commutation rules given in (4.6). Proceeding in the same way that we did for a spin-1 particle, we obtain for a particle of four-momentum p and spin 2, the sum over polarization given in Eq. (4.10).

$$S^{\mu\nu\mu'\nu'}(p) = \sum_{\lambda=1}^{2}\epsilon^{\mu\nu}(p,\lambda)\epsilon^{*\mu'\nu'}(p,\lambda) = \tfrac{1}{8}\{X^{\nu\nu'}(p)Y^{\mu\mu'}(p) - X^{\mu\mu'}(p)Y^{\nu\nu'}(p) + X^{\nu\mu'}(p)Y^{\nu'\mu}(p) + X^{\nu'\mu}(p)Y^{\nu\mu'}(p) - 2X^{\mu\nu}(p)X^{\mu'\nu'}(p)\}, \qquad (4.10)$$

where $X^{\mu\nu}(p)$ is given by (4.7) and $Y^{\mu\nu}(p)$ is defined in Table II ($\mu,\nu,\mu',\nu' = 0,1,2,3$).

As for a spin-1 particle, we derived the relations satisfied by $\epsilon^{\mu\nu}(p,\lambda)$ from the conditions imposed to $f^{\mu\nu}$ and given in Table I. One has

$$\epsilon^{00}(p,\lambda) = \epsilon^{0\mu}(p,\lambda) = \epsilon^{\mu 0}(p,\lambda), \quad \epsilon^{\mu}{}_{\mu}(p,\lambda) = 0,$$

$$\epsilon^{\mu\nu}(p,\lambda) = \epsilon^{\nu\mu}(p,\lambda), \quad p^{\mu}\epsilon_{\mu\nu}(p,\lambda) = 0,$$

$$ip^0\epsilon^{\mu\nu}(p,\lambda) - i\epsilon^{0\nu\alpha\beta}p^{\alpha}\epsilon^{\mu\beta}(p,\lambda) = 0.$$

Concluding this section, we remark that we have obtained with our formalism all usual results. In the next section we will give some effective interactions as stated in Sec. III.

V. EFFECTIVE INTERACTIONS

In order to classify the interactions, we remember that under Poincare coordinate transformation, a $2s$ rank spinor transform as:

$$\psi(x) \to \psi'(x') = D(l)\otimes\cdots\otimes D(l)\psi(x),$$

where the operator $D(l)$ satisfies

$$D^{-1}(l)\gamma^{\mu}D(l) = l^{\mu}{}_{\nu}\gamma^{\nu}$$

and

$$\gamma^0 D^{+}\gamma^0 = D^{-1}.$$

The $2s$ rank spinor $\bar{\psi}(x)$ defined by

$$\bar{\psi}(x) = \psi^{+}(x)\gamma^0\otimes\cdots\otimes\gamma^0$$

transforms like

$$\bar{\psi}(x) \to \bar{\psi}'(x') = \bar{\psi}(x)D^{-1}(l)\otimes\cdots\otimes D^{-1}(l).$$

Note that all tensor products written above contains $2s$ terms. Since BW's theory is a generalization of Dirac's theory, we can use the well known results for Dirac's theory in order to ensure ourselves that the following forms transform as

(i) scalar $\quad 1 = \frac{1}{2s}(I\otimes\cdots\otimes I + \cdots + I\otimes\cdots\otimes I),$

(ii) pseudoscalar $\quad \Gamma^5 = \frac{1}{2s}(\gamma^5\otimes\cdots\otimes T + \cdots + I\otimes\cdots\otimes\gamma^5),$

(iii) vector $\quad \Gamma^{\mu} = \frac{1}{2s}(\gamma^{\mu}\otimes\cdots\otimes I + \cdots + I\otimes\cdots\otimes\gamma^{\mu}),$

(iv) pseudovector $\quad \Gamma^{\mu}\Gamma^5 = \frac{1}{2s}(\gamma^{\mu}\gamma^5\otimes\cdots\otimes I + \cdots + I\otimes\cdots\otimes\gamma^{\mu}\gamma^5),$

(v) tensor $\quad \Sigma^{\mu\nu} = \frac{1}{2s}(\sigma^{\mu\nu}\otimes\cdots\otimes I + \cdots + I\otimes\cdots\otimes\sigma^{\mu\nu}),$

where all the above terms contains $2s$ terms, composed of $2s$ tensor products.

We denote by Γ^n the following set of matrices:

$$\Gamma^n = \{1,\Gamma^5,\Gamma^{\mu},\Gamma^{\mu}\Gamma^5,\Sigma^{\mu\nu}\}. \qquad (5.1)$$

We are now able to construct Dirac forms for different particles with the same spin, such as

$$\bar{\psi}\Gamma^n\psi' = \{\bar{\psi}\psi',\bar{\psi}\Gamma^5\psi',\bar{\psi}\Gamma^{\mu}\psi',\bar{\psi}\Gamma^{\mu}\Gamma^5\psi',\bar{\psi}\Sigma^{\mu\nu}\psi'\}, \qquad (5.2)$$

where in (5.2) ψ' (and/or $\bar{\psi}$) are $2s$-rank spinors describing different particles with or without mass. In fact we constructed the above forms involving particles with same spin particles, i.e., forms that contain:

massive particles–massive particles[2], (5.3a)

massive particles–massless particles, (5.3b)

massless particles–massless particles. (5.3c)

We can also construct Dirac forms mixing particles with different spins. For convenience we will use, from now on, the following notation:

Θ_a ($\tilde{\Theta}_a$) is a massive (massless) particle of spin $\frac{1}{2}$;

$\zeta_{a_1a_2}$ ($\tilde{\zeta}_{a_1a_2}$) is a massive (massless) particle of spin 1;

$\phi_{a_1a_2a_3}$ ($\tilde{\phi}_{a_1a_2a_3}$) is a massive (massless) particle of spin $\frac{3}{2}$;

$\xi_{a_1a_2a_3a_4}$ ($\tilde{\xi}_{a_1a_2a_3a_4}$) is a massive (massless) particle of spin 2, where tilde means different particles with the same spin.

All Dirac forms involving particles with different spin are listed in Table III. In this table all considered particles are indifferently massive or massless and all expressions are supposed to be symmetrized in spin indices. We omitted contracted spin indices.

We can now construct interactions involving: (i) particles with the same spin whether they are massive or massless for example scalar-type interactions like as $(\bar{\psi}\Gamma^n\psi')(\bar{\psi}\Gamma_n\psi')$, where here Γ^n denotes the same matrix in the two factors; vector-type interactions like as

$$(\bar{\psi}\Gamma^{\mu}\psi');\bar{\psi}\Sigma^{\mu\nu}\psi' q_\nu$$

where q is the momentum transfer, etc., and
(ii) particles with different spins

In this case Table III gives us some interactions. Obvisously since Table III is not exaustive, we can also obtain, for example, vector-type interactions in the following manner for particles with or without mass:

$$\{\bar{\xi}\Sigma^{\mu\nu}\phi\Theta q_\nu,\bar{\xi}\Sigma^{\mu\nu}\zeta\zeta' q_\nu,\bar{\xi}\Sigma^{\mu\nu}\zeta\Theta\Theta' q_\nu,\bar{\xi}\Sigma^{\mu\nu}\Theta\Theta'\Theta'' q_\nu\};$$
$$\{\bar{\phi}\Sigma^{\mu\nu}\zeta\Theta q_\nu,\bar{\phi}\Sigma^{\mu\nu}\Theta\Theta'\Theta'' q_\nu\};$$
$$\{\bar{\zeta}\Sigma^{\mu\nu}\Theta\Theta' q_\nu\},$$

q being, in all these cases, the momentum transfer. An important property of the last interaction is that the currents so constructed are conserved due to the antisymmetry of the tensor $\Sigma^{\mu\nu}$. We obtain the usual formulation using the decomposition of BW's fields for massive[2] or massless particles. We will develop explicit interactions and study their phenomenological properties in a future publication.

VI. CONCLUSIONS

The aim of this paper is to give the theoretical principles necessary to construct all types of interactions between massive (or massless) fields. Taking the interactions of Sec. V as a starting point, we deduce Feynman rules for massive or massless particles with spin $\frac{1}{2}$, 1, $\frac{3}{2}$, and 2. In addition we obtained another interesting result. We describe free particles with spin 1 and 2 in a quite different manner than usual.[4] In order to connect our theory with quantum electrodynamics we will, based on our formulation, construct Lagrangians of interaction involving photons and spin-$\frac{1}{2}$ particles.[7] Later we will generalize this treatment of QED to particles with higher spin.

TABLE III. Dirac forms for particles with different spin.

Spin s = something	Dirac forms
2	$\bar{\xi}\Gamma^n\Phi\Theta$
	$\bar{\xi}\Gamma^n\zeta\zeta'$
	$\bar{\xi}\Gamma^n\zeta\Theta\Theta'$
	$\bar{\xi}\Gamma^n\Theta\Theta'\Theta\Theta''\Theta'''$
$\frac{3}{2}$	$\bar{\Phi}\Gamma^n\zeta\Theta$
	$\bar{\Phi}\Gamma^n\Theta\Theta'\Theta''$
1	$\bar{\zeta}\Gamma^n\Theta\Theta'$

ACKNOWLEDGMENTS

We are particularly indebted to Professor G. da Costa Marques for his interest and discussions, as well as for critical reading of the manuscript. We would like to thank also Professor O. J. P. Eboli for interesting discussions and Professor C. A. Pereira in the preparation of the manuscript. We would like to thank the Instituto de Fisica Teorica, UNESP, and the Departmento de Fisica Matematica, USP for the kind hospitality.

We would like to thank also the Fundação de Amparo a Pesquisa do Estado de Sao Paulo (FAPESP) for their partial financial support.

APPENDIX

We used in this paper the metric $g_{\mu\nu} = (1, -1, -1, -1)$. The Pauli matrices are defined as below

$$\sigma^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \quad \sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix};$$
$$\sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}; \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

By convention $\sigma^\mu = (\sigma^0 = \mathbf{1},\boldsymbol{\sigma})$ designates $(\sigma^\mu)_{b_1\dot{b}_2}$ $b_1,b_2 = 1,2$. We have used also the property that

$$(\sigma^0)_{b_1\dot{b}_2} = (\sigma_0)_{b_1\dot{b}_2}; \quad (\sigma^k)_{b_1\dot{b}_2} = -(\sigma_k)_{b_1\dot{b}_2},$$

where C_1 denotes the charge conjuction matrix in (2×2) space and obeys

$$^tC_1 = -C_1; \quad C_1^{\ +} = C_1^{\ -1}.$$

Useful results for chiral representaion are: (i) Dirac matrices

$$\gamma^\mu = (\gamma^0,\boldsymbol{\gamma}); \quad \gamma^\mu = \begin{pmatrix} 0 & -\sigma_\mu \\ -\sigma^\mu & 0 \end{pmatrix};$$

(ii) Charge conjugation matrix

$$C = \begin{pmatrix} C_1 & 0 \\ 0 & C_1^{-1} \end{pmatrix}; \quad \gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix};$$
$$\boldsymbol{\Sigma} = \gamma^5\gamma^0\boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\sigma} & 0 \\ 0 & \boldsymbol{\sigma} \end{pmatrix}; \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix};$$
$$\sigma^{\mu\nu} = \tfrac{i}{2}[\gamma^\mu,\gamma^\nu].$$

[1]V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. (USA) **34**, 211 (1948).
[2]D. Spehler and V. Pleitez, "Effective theories for particles of spin $\frac{1}{2}$, 1, $\frac{3}{2}$, and 2: I. The massive case" (to appear).
[3]D. Spehler, Ph.D. thesis, Preprint CRN-HE 79-10, Strasbourg, 1979.
[4]H. Van Dam and M. Veltmann, Nucl. Phys. B **22**, 402 (1970). S. Weinberg, Phys. Rev. B **133**, 1318 (1964).
[5]J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965).
[6]J. Leite Lopes and D. Spehler, Lett. Nuovo Cimento **25**, 101 (1979).
[7]D. Spehler and G. Narques, "Left right asymmetry and minimal coupling" (to appear).

On the BRS's

Luca Lusanna
INFN, Sezione di Firenze, Largo E. Fermi 2, Arcetri, 50127 Firenze, Italy

(Received 19 July 1989; accepted for publication 30 August 1989)

After an analysis of the concept of Lagrangian gauge fixing, it is shown that the arbitrariness in the parametrization of gauge transformations gives rise to a whole family of classical BRS transformations. This is explicitly shown for the free-relativistic particle. Two inequivalent classes of BRS Lagrangians are defined. While the former generates a Kato–Ogawa-like Hamiltonian BRS formalism, the latter gives rise to the Batalin–Fradkin–Vilkovisky theory. A comparison is made between these Hamiltonian theories, the multitemporal description of 1st-class constraints, the Konstant–Sternberg and Loll approaches, and the Bonora–Cotta Ramusino interpretation of ghosts. The relevance of an equivariance condition for the BRS observables is shown. The quantum BRS theory is briefly discussed.

I. INTRODUCTION

Looking at the literature concerning the theory and the applications of the BRS methodology, one remains astonished at how many variants exist and, on the other hand, by the nearly total absence of attempts to extract a unified picture. Therefore, we felt the need to investigate this matter, starting from the classical level with a finite number of degrees of freedom. To simplify the subject even further, we shall consider the example of the free relativistic particle as a guide in the study of the Lagrangian and Hamiltonian forms of the BRS method. Only in Secs. IV and V shall we investigate systems whose constraints are associated to a Lie algebra, to make contact with more sophisticated concepts of the BRS theory.

Let us first make some general considerations on the problem of the Lagrangian gauge fixings. If we have a singular Lagrangian $L_0(q,\dot q)$ on a configuration space Q with local coordinates q^i, $i=1,\dots,n$, whose Hessian matrix $A_{0ij} = \partial^2 L_0/\dot q^j \partial\dot q^i$ has only one null eigenvalue, for the sake of simplicity, then there exists[1,2] a set of Nœther transformations $\delta_0 q^i = \Sigma_{j=0}^J \epsilon^{(j)}(t)\xi^i_{J-j}(q,\dot q)$ for some $J[\epsilon(t)$ is an arbitrary function$]$ such that $\delta_0 L_0 = dF_0/dt$. Since the $\delta_0 q^i$ are gauge transformations, in the phase space $T^*Q(p_0 = \partial L_0/\partial\dot q^i)$ there will be a set of 1st-class constraints $\overline{G}_{J-j}(q,p_0)\approx 0, j=0,1,\dots,J$ ($\overline{G}_0$ is the primary one), so that $\delta_0 q^i = \{q^i,\overline{G}\}$ with $\overline{G} = \Sigma_{j=0}^J \epsilon^{(j)}(t)\overline{G}_{J-j}$. At the Lagrangian level we obtain $\overline{G}_{J-j}(q,p_0) = G_{J-j}(q,\dot q) \doteq 0$, because they are the acceleration-independent subset of the Euler–Lagrange equations ($\doteq$ means evaluated on the solutions of the Euler–Lagrange equations).

It is well known that to fix the gauge in T^*Q one adds a set of Dirac gauge-fixing constraints $\overline{\chi}_{J-j}(q,p_0)\approx 0$, $j=0,1,\dots,J$, such that $\det|\{\overline{G}_h,\overline{\chi}_k\}|\neq 0$, builds the corresponding Dirac brackets and so obtains a copy of the reduced phase space.

To have a consistent model, one must assume that L_0 is such that the commutators of any order of the single gauge transformations $\delta_{0j}q^i = \epsilon^{(j)}(t)\xi^i_{J-j}, j=0,1,\dots,J$, must close upon the gauge transformations themselves[1] (gauge algebra hypothesis; the algebra is open when the $\delta_0 q^i$ are velocity dependent) and the same must happen for the commutator of the gauge transformations with the deterministic evolution implied by L_0. In T^*Q this is the Dirac test $\{\overline{G}_h,\overline{G}_k\} = \overline{C}^r_{hk}(q,p_0)\overline{G}_r$, $\{\overline{H},\overline{G}_h\} = \overline{V}^k_h(q,p_0)\overline{G}_k$, where $\overline{H}$ is the canonical Hamiltonian.

To introduce a Lagrangian gauge fixing, one defines the new Lagrangian:

$$L(q,\dot q) = L_0(q,\dot q) + \tfrac{1}{2}\alpha[\psi(q,\dot q)]^2, \tag{1}$$

with the only condition $\det(A_{ij} = A_{0ij} + \alpha A_{1ij})\neq 0$, which ensures that the original gauge invariance has been broken: $\delta_0 L = dF_0/dt + \alpha\psi\,\delta_0\psi$. When $\delta_0\psi\neq 0$, $\delta_0\psi\neq\psi^{-1}(dF_1/dt)$, no gauge freedom is left and, as is shown in Ref. 2, $\psi(q,\dot q)$ may be obtained by prefixing the gauge, i.e., by assigning $\lambda^0(t) = \rho_\psi(q,\dot q)$ in the Dirac Hamiltonian $\overline{H}_D = \overline{H} + \lambda^0(t)\overline{G}_0$ and by making a suitable Legendre transformation. To assign $\lambda^0(t) = \bar\rho_\psi(q,p_0) = \rho_\psi(q,\dot q)$ does not mean choosing Dirac gauge-fixing constraints, but only restricting the gauge fixing $\overline{\chi}_0\approx 0$ for the primary 1st-class constraint $\overline{G}_0\approx 0$ to be a solution of the differential equation $\dot{\overline{\chi}}_0 \doteq \{\overline{\chi}_0,\overline{H}_D\} = \{\overline{\chi}_0,\overline{H}\} + \bar\rho_\psi(q,p_0)\{\overline{\chi}_0,\overline{G}_0\}\approx 0$. The other gauge fixings $\overline{\chi}_{J-j}, j=0,1,\dots,J-1$, then have to satisfy certain compatability conditions shown in Ref. 2 for each chosen solution $\overline{\chi}_0$. Since, in general, a global gauge fixing is not allowed (think the Gribov ambiguity in field theory), it becomes interesting to see whether with a $\psi(q,\dot q)$ not in the previous class we can give some meaning to the equation $\delta_0\psi = 0$ (we do not consider in this paper the more general condition $\delta_0\psi = \psi^{-1}\,dF_1/dt$ for some F_1); $\delta_0\psi = 0$ cannot be satisfied with $\delta_0 q^i$ depending on an arbitrary $\epsilon(t)$, because L is regular. However, we could read $\delta_0\psi = (\partial\psi/\partial q^i)\delta_0 q^i + (\partial\psi/\partial\dot q^i)\delta_0\dot q^i = 0$ as an equation for $\epsilon(t)$ and look for solutions $\tilde\epsilon_\psi(t)$ such that $\tilde\delta_0\psi = 0$ for $\tilde\delta_0 q^i = \delta_0 q^i|_{\epsilon=\tilde\epsilon}$. In general, we may have solutions $\tilde\epsilon_\psi(t|q,\dot q)$, but for the sake of simplicity we shall consider only ψ such that $\tilde\epsilon_\psi = \tilde\epsilon_\psi(t)$. Therefore, if we restrict the

 0022-2488/90/020428-18$03.00

arbitrary function $\epsilon(t)$ to the special class of functions $\tilde{\epsilon}_\omega(t)$, i.e., if we restrict ourselves to the subset $\tilde{\delta}_0 q^i$ of the gauge transformations $\delta_0 q^i$, we get $\tilde{\delta}_0 L = \tilde{\delta}_0 L_0 = dF_0/dt$, which is a quasiinvariance. Therefore, the $\tilde{\delta}_0 q^i$ describe the "residual gauge freedom" left by the Lagrangian gauge fixing ψ.

Let us remark that if in Eq. (1) we had put a Lagrangian multiplier $\alpha(t)$, so that $(\psi(q,\dot{q}))^2 = 0$ results as an equation of motion, we would have had problems with the Jacobi equations (see Ref. 2). Therefore, this is not a good way to impose a Lagrangian gauge fixing.

Since L is regular, $\tilde{\delta}_0 L = dF_0/dt$ will imply the existence of as many constants of the motion as there are independent time derivatives of the special functions $\tilde{\epsilon}_\psi(t)$. Therefore, we can write $\tilde{\delta}_0 q^i = \Sigma_{j=0}^{J_\psi} \tilde{\epsilon}_\psi^{(j)}(t)\{q^i, \overline{G}_{J-j}(q,p)\}$ with $J_\psi < J$. Here, $p_i = \partial L/\partial \dot{q}^i = p_{0i} + \alpha p_{1i}$, $p_{1i} = \psi\, \partial\psi/\partial\dot{q}^i$, so that $\overline{G}_{J-j}(q, p = p_0 + \alpha p_1) \mapsto \overline{G}_{J-j}(q,p_0)$ for $\alpha \to 0$, $j \leqslant J_\psi$. If $L_{0i} = \partial L_0/\partial q^i - (d/dt)(\partial L_0/\partial \dot{q}^i) \doteq 0$ are the Euler–Lagrange equations for L_0, those for L are

$$L_i = L_{0i} + \alpha\left[\psi\frac{\partial\psi}{\partial q^i} - \frac{d}{dt}\left(\psi\frac{\partial\psi}{\partial\dot{q}^i}\right)\right]. \tag{2}$$

If we put $\widetilde{G}_{J-j}(q,p) = \widetilde{G}_{J-j}(q,\dot{q}), j \leqslant J_\psi$, then $\tilde{\delta}_0 L = dF_0/dt$ implies $\widetilde{G}_{J-j} \doteq 0$ by using $L_i \doteq 0$. Since, due to the definition of gauge transformations,[1,2] we have $\delta_0 L_{0i} = J_{0i}(\delta_0 q^k) \doteq 0$ for $L_{0i} \doteq 0$, where J_{0i} are the Jacobi equations, we get

$$\begin{aligned} J_i(\tilde{\delta}_0 q^i) &= \tilde{\delta}_0 L_i = \tilde{\delta}_0 L_{0i} + \alpha\left[\psi\tilde{\delta}_0\left(\frac{\partial\psi}{\partial q^i} - \frac{d}{dt}\frac{\partial\psi}{\partial\dot{q}^i}\right) - \dot{\psi}\tilde{\delta}_0\frac{\partial\psi}{\partial\dot{q}^i}\right] \\ &\underset{L_{0i}\doteq 0}{=} \alpha\left[\psi\tilde{\delta}_0\left(\frac{\partial\psi}{\partial q^i} - \frac{d}{dt}\frac{\partial\psi}{\partial\dot{q}^i}\right) - \dot{\psi}\tilde{\delta}_0\frac{\partial\psi}{\partial\dot{q}^i}\right]\Bigg|_{L_{0i}\doteq 0}. \end{aligned} \tag{3}$$

The constants of the motion $\widetilde{G}_{J-j}(q,\dot{q}), j \leqslant J_\psi$, are rather special: (i) they are associated with Nœther transformations $\tilde{\delta}_0 q^i$ depending on the functions $\tilde{\epsilon}_\psi(t)$; (ii) they satisfy $\widetilde{G}_{J-j}(q,\dot{q}) \mapsto G_{J-j}(q,\dot{q}) \doteq 0$ for $\alpha \mapsto 0$ due to $L_{0i} \doteq 0$. Moreover, on the solutions of the equation $\psi(q,\dot{q}) = 0$ we have $p_i = p_{0i}$, $L_i = L_{0i}$, $J_i(\tilde{\delta}_0 q^k) \doteq 0$ for $L_{0i} \doteq 0$, and therefore $\widetilde{G}_{J-j} = G_{J-j} \doteq 0$ for $L_{0i} \doteq 0$. This implies that $\psi(q,\dot{q}) = 0$ is compatible with $L_{0i} \doteq 0$: among the solutions of $L_{0i} \doteq 0$ there is a special family which satisfies $L_{0i} \doteq 0$, $\psi = 0$, so that $\psi(q,\dot{q}) = 0$ is a restriction of the gauge freedom in the solutions of $L_{0i} \doteq 0$, which still leaves a residual gauge freedom. In this special family of solutions we have $\widetilde{G}_{J-j} = G_{J-j} \doteq 0, j \leqslant J_\psi$, and this explains why in this case the first Nœther theorem works with the Nœther transformations $\tilde{\delta}_0 q^i$ depending on the special class of functions $\tilde{\epsilon}_\psi(t)$ rather than on constant parameters.

When we select this special family of extremals of L, defined by $\psi = 0$, by means of a restriction of the initial data, we obtain that class of extremals of L_0 which is consistent with the restriction of the original gauge freedom to the residual one described by the $\tilde{\delta}_0 q^i$ with $\tilde{\epsilon}_\psi(t)$. This means that the originally arbitrary velocity function (Ref. 1) $g_\lambda^0(q,\dot{q})$ of L_0 has its arbitrariness consistently restricted with the $\tilde{\delta}_0 q^i$'s. Here g_λ^0 is that function, nonprojectable to phase space, which exists because the singular nature of L_0 forbids us from expressing all the velocities as functions of q^i, p_{0i} by means of $p_{0i} = \partial L_0/\partial\dot{q}^i$ (Refs. 1 and 2): its functional form is chosen in such a way that the Hamilton equations imply $g_\lambda^0(q,\dot{q}) \doteq \lambda^0(t)$. Therefore, the restriction to the residual gauge freedom implies that the arbitrary Dirac multiplier is also restricted to a special class of functions $\tilde{\lambda}_\psi^0(t)$. Let us remark that $\psi = 0$ could uniquely determine $g_\lambda^0(q,\dot{q}) \doteq \lambda^0(\tau) = \tilde{\lambda}_\psi^0$: in this case, the Lagrangian residual gauge freedom becomes the Hamiltonian freedom of choosing anyone of the Dirac gauge fixings compatible with the given Dirac multiplier (think to the residual conformal gauge transformations of the string in the orthonormal gauge).

Let us also note that $L' = L_0 + \alpha(t)\psi$ does not work as a Lagrangian gauge fixing, because even if the Euler–Lagrange equation for $\alpha(t)$ implies $\psi \doteq 0$, the other Euler–Lagrange equations do not reduce to $L_{0i} \doteq 0$ when $\psi \doteq 0$. Instead, the Lagrangian $L'' = L_0 + b\psi - b^2/2$, where $b(t)$ is a nonlinear Lagrange multiplier, is relevant for our problem because its Euler-Lagrange equation $L_b'' = \psi - b \doteq 0$ has the solution $b \doteq \psi$, $b \neq 0$: If we put this solution in L'', we recover L of Eq. (1) with $\alpha = 1$. Here, $L_b'' \doteq 0$ also has the following second sector of solutions: $b \doteq 0, \psi \doteq 0$. In this sector, the other Euler–Lagrange equations for L'' coincide with those for L_0 and the net effect is now to restrict the extremals of L_0 to the class satisfying $\psi = 0$. Therefore, the use of L'' allows us to separate the two classes of extremals of the L of Eq. (1). The phase space description associated with L'' has either two sectors with different constraints (this is an example of ramification of sectors and proliferation of constraints[2] as we shall see in Sec. II) when $\psi(q,\dot{q})$ is projectable to $\bar{\psi}(q,p)$ or only one sector with well-defined constraints when $\psi(q,\dot{q})$ is not projectable. In this last case, the two Lagrangian sectors correspond to two sectors of the Hamilton equations.

This ends our discussion of the Lagrangian gauge-fixing problem, which was modeled on the well-known methods used for the electromagnetic and Yang–Mills fields. However, the restriction $\epsilon(t) \mapsto \tilde{\epsilon}_\psi(t)$ does not assure us that the subset of gauge transformations $\tilde{\delta}_0 q^i$ satisfies the conditions for being a consistent gauge subalgebra of the original one. This is why one has to introduce the BRS method[3] in which $\tilde{\epsilon}_\psi(t) \mapsto \rho c(t)$ with $c(t)$ the ghost and ρ a constant odd parameter: The special Nœther transformations $\tilde{\delta}_0 q^i$ become the BRS transformations $\delta_s q^i$. As we shall see in Sec. IV, the nilpotency condition $\delta_s^2 = 0$ is just the condition for having a global gauge algebra. Then the first Nœther theorem, which associates the constant of the motion $\Sigma_{j=0}^{J} \tilde{\epsilon}_\psi^{(j)}(t)\widetilde{G}_{J-j}$ to $\tilde{\delta}_0 q^i$, now determines the BRS charge $\widetilde{\Omega}$ from the BRS symmetry of the BRS Lagrangian under the $\delta_s q^i$'s.

II. THE CLASSICAL FREE-RELATIVISTIC PARTICLE

Instead of trying to treat the general case, let us use the free-relativistic particle as a simple example whereby to study the various variants of the BRS method. We shall first consider the following Lagrangian containing the einbein

$\lambda = \epsilon^{\omega} \neq 0$:

$$L = \dot{x}^2/4\lambda + m^2\lambda, \tag{4}$$

and only at the end shall we speak about $L_{\pm} = \pm m\sqrt{\dot{x}^2}$. Here, L can be obtained as the Dirac Lagrangian associated with $L_{\pm}$, if we make the inverse Legendre transformation of the Dirac Hamiltonian $\bar{H}_{D\pm} = \lambda_{\pm}(p^2 - m^2)$ without using $p^2 - m^2 \approx 0$: the einbein $\lambda(t)$ is the Dirac multiplier and L takes into account simultaneously the two branches of the mass hyperboloid $p^2 = m^2$, which instead are treated separately with $L_{\pm}$: $\lambda_+ > 0$ implies $p_0 = +\sqrt{\mathbf{p}^2 + m^2}$, $\lambda_- < 0$ the other branch. The momenta and the Euler–Lagrange equations for L are, respectively,

$$p_\mu = \frac{\partial L}{\partial \dot{x}^\mu} = \frac{\dot{x}_\mu}{2\lambda}, \qquad \pi_\lambda = \frac{\partial L}{\partial \dot{\lambda}} = 0, \tag{5}$$

$$L_\mu = -\frac{d}{d\tau}\frac{\dot{x}_\mu}{2\lambda} = -\frac{1}{2\lambda}\left(\ddot{x}_\mu - \frac{\dot{\lambda}}{\lambda}\dot{x}_\mu\right) \doteq 0, \qquad L_\lambda = m^2 - \dot{x}^2/4\lambda^2 \doteq 0. \tag{6}$$

The solutions of Eqs. (6) are

$\lambda(t)$ arbitrary, $p^\mu(t) \doteq mA^\mu$,

$$x^\mu(\tau) \doteq x^\mu(0) + 2mA^\mu \int_0^\tau d\tau_1\, \lambda(\tau_1), \quad A^2 = 1, \tag{7}$$

where A^μ is a constant timelike unit vector. The Hessian matrix of L has a null eigenvalue, so that under the following Nœther transformations:

$$\delta\lambda = \frac{d}{d\tau}(\epsilon(\tau)f), \qquad f = f(\dot{x}^2, \lambda, \dot{\lambda}), \qquad \delta x^\mu = \epsilon(\tau)(\dot{x}^\mu/\lambda)f, \tag{8}$$

we get the following quasiinvariance:

$$\delta L = \frac{d}{d\tau}\left(\epsilon(\tau)f\frac{L}{\lambda}\right). \tag{9}$$

Here $\epsilon(\tau)$ is an arbitrary function of τ and f is an arbitrary function of its arguments (to have Poincarè invariance we excluded a dependence on x^μ). The arbitrariness of f reflects the arbitrariness in the parametrization of the gauge transformations: $f = 1$ amounts to using an orthonormal set of eigenvectors of the Hessian.[1,2] This arbitrariness is also present in the first Nœther theorem (see its extensions[4]): there is an infinite family of Nœther transformations associated with each constant of the motion.[4,1,2] The resulting Nœther identities[1,2] are

$$\begin{aligned}
&f\pi_\lambda \equiv 0,\\
&f(p^2 - m^2) + 2\dot{f}\pi_\lambda + f\dot{\pi}_\lambda \equiv -fL_\lambda,\\
&\frac{d}{d\tau}[f(p^2 - m^2) + \dot{f}\pi_\lambda] \equiv -(f(\dot{x}^\mu L_\mu/\lambda) + \dot{f}L_\lambda),\\
&\Downarrow\\
&\pi_\lambda \equiv 0,\\
&0 \equiv \dot{\pi}_\lambda \equiv -(p^2 - m^2) - L_\lambda,\\
&0 \doteq -\frac{dL_\lambda}{d\tau} \equiv \frac{d}{d\tau}(p^2 - m^2) \equiv -\frac{\dot{x}^\mu L_\mu}{\lambda},
\end{aligned}
\quad \Rightarrow \begin{cases} p^2 - m^2 \equiv -L_\lambda \doteq 0,\\ \dot{x}^\mu L_\mu - \lambda\dfrac{dL_\lambda}{d\tau} \equiv 0. \end{cases} \tag{10}$$

They reproduce the Dirac–Bergmann algorithm: in phase space $\pi_\lambda \approx 0$ is the primary 1st-class constraint, $p^2 - m^2 \approx 0$ the secondary 1st-class one, and the Dirac Hamiltonian is $\bar{H}_D = \lambda(p^2 - m^2) + \mu(\tau)\pi_\lambda$ [$\mu(\tau)$ is the Dirac multiplier]. Since the constraints are 1st class, there is the contracted Bianchi identity $\dot{x}^\mu L_\mu - \lambda(dL_\lambda/d\tau) \equiv 0$ and the arbitrary velocity function of L (nonprojectable to phase space) is $\dot{\lambda} \doteq \mu(\tau)$. The canonical transformation[5]

$$\begin{matrix} \lambda & x^\mu \\ \pi_\lambda & p_\mu \end{matrix} \mapsto \begin{matrix} \lambda & \dfrac{p\cdot x}{\eta\sqrt{p^2}} & \mathbf{z} = \dfrac{\eta\sqrt{p^2}}{m}\left(\mathbf{x} - \dfrac{\mathbf{p}}{p^0}x^0\right), \\ \pi_\lambda & \eta\sqrt{p^2} - m & \mathbf{k} = \dfrac{m}{\eta\sqrt{p^2}}\mathbf{p} \quad \eta = \pm, \end{matrix} \tag{11}$$

exhibits a canonical basis in which two of the momenta are the constraints, λ and $p\cdot x/\eta\sqrt{p^2}$ are the gauge variables and $\mathbf{z}$, $\mathbf{k}$ are strong observables. Since, besides $\eta\sqrt{p^2} - m = \eta\sqrt{\dot{x}^2}/2\lambda - m$ and $p\cdot x/\eta\sqrt{p^2} = \eta\dot{x}\cdot x/\sqrt{\dot{x}^2}$, we have

$$\mathbf{z} = \frac{\eta\sqrt{\dot{x}^2}}{zm\lambda}\left(\mathbf{x} - \frac{\dot{\mathbf{x}}}{\dot{x}^0}x^0\right) \doteq \eta\left[\mathbf{x}(0) - \frac{\mathbf{A}}{A^0}x^0(0)\right], \qquad \mathbf{k} = m\dot{\mathbf{x}}/\eta\sqrt{\dot{x}^2} \doteq \eta m\mathbf{A}. \tag{12}$$

We see that at the Lagrangian level, these strong observables correspond to a possible parametrization of the independent initial data. Indeed, the only physical relevant quantities are $\mathbf{x}(x^0) \doteq \mathbf{x}(x^0 = 0) + (\mathbf{A}/A^0)x^0$ with $\mathbf{x}(x^0 = 0) = \mathbf{x}(\tau = 0) - (\mathbf{A}/A^0)x^0(\tau = 0)$ and $(\mathbf{A}/A^0) \doteq [d\mathbf{x}(x^0)]/dx^0$.

Let us now put $\epsilon(\tau) = \rho c(\tau)$, where $c(\tau)$ is the ghost. Then Eqs. (9) and (10) become

$$\delta_s\lambda = \dot{f}c + f\dot{c}, \qquad \delta_s x^\mu = f(\dot{x}^\mu/\lambda)c, \quad f = f(\dot{x}^2, \lambda, \dot{\lambda}), \qquad \delta_s L = \frac{d}{d\tau}\left(f\frac{L}{\lambda}c\right), \tag{13}$$

Since $\partial f/\partial \dot{x}^\mu = 2\dot{x}_\mu\, \partial f/\partial \dot{x}^2$, we get

$$\delta_s f = 2\frac{\partial f}{\partial \dot{x}^2}\frac{\dot{x}_\mu}{\lambda}\left[\left(f\ddot{x}^\mu + \left(\dot{f} - \frac{\dot{\lambda}}{\lambda}f\right)\dot{x}^\mu\right)c + f\dot{x}^\mu\dot{c}\right] + \frac{\partial f}{\partial \lambda}(\dot{f}c + f\dot{c}) + \frac{\partial f}{\partial \dot{\lambda}}(\ddot{f}c + 2\dot{f}\dot{c} + f\ddot{c}). \tag{14}$$

If we evaluate $\delta_s^2\lambda$ and $\delta_s^2 x^\mu$ we get the following result:

$$\begin{matrix}\delta_s^2\lambda = 0 \\ \delta_s^2 x^\mu = 0\end{matrix} \Rightarrow \delta_s c = \left(\frac{2\dot{x}^2}{\lambda}\frac{\partial f}{\partial \dot{x}^2} + \frac{\partial f}{\partial \lambda} + 2\frac{\dot{f}}{f}\frac{\partial f}{\partial \dot{\lambda}}\right)c\dot{c} + \frac{\partial f}{\partial \dot{\lambda}}c\ddot{c}, \tag{15}$$

which is the BRS transformation rule of the ghost c when the gauge transformations are parametrized with f. To remember this fact it would be more correct to use the symbol δ_s^f, but to abbreviate the notation we shall go on with δ_s. Next we get

$$\delta_s^2 c = -\frac{2}{f}\frac{\partial f}{\partial \dot{\lambda}}\frac{d}{d\tau}\left(f\frac{\partial f}{\partial \dot{\lambda}}\right)c\dot{c}\ddot{c},$$
$$\delta_s^2 c = 0 \Rightarrow \frac{\partial f}{\partial \dot{\lambda}} = 0 \quad \text{or} \quad \frac{d}{d\tau}\ln\left(\frac{\partial f^2}{\partial \dot{\lambda}}\right) = 0. \tag{16}$$

Our first result is that not all the parametrizations f produce a nilpotent δ_s^f (i.e., consistent with the hypothesis of gauge algebra), but only those for which

$$\text{Either } f = f(\dot{x}^2,\lambda) \Rightarrow \delta_s c = \left(\frac{2\dot{x}^2}{\lambda}\frac{\partial f}{\partial \dot{x}^2} + \frac{\partial f}{\partial \lambda}\right)c\dot{c},$$

$$\text{Or } f^2 = a\dot{\lambda} + g(\dot{x}^2,\lambda) \Rightarrow \delta_s c = \frac{1}{2f}\left[\left(\frac{2\dot{x}^2}{\lambda}\frac{\partial g}{\partial \dot{x}^2} + \frac{\partial g}{\partial \lambda} + 2a(\dot{f}/f)\right)c\dot{c} + ac\ddot{c}\right]. \tag{17}$$

In what follows we shall, for simplicity's sake, only consider the case $f = f(\dot{x}^2,\lambda)$. Let us note that if in the class of the allowed f we make the change of parametrization $f_2 = \dot{f}f_1$ then we get the following rule connecting the ghosts c_2 and c_1 associated with the parametrizations f_2, f_1, respectively: $c_1 = fc_2$.

Before going on, let us consider the BRS transformation properties of the Lagrangian counterparts of the quantities defined in Eqs. (11) and (12). Since $\pi_\lambda \equiv 0$, we have, from Eqs. (13),

$$\delta_s\lambda = \dot{f}c + f\dot{c},$$

$$\delta_s(\eta\sqrt{p^2} - m) = \delta_s\left(\frac{\eta\sqrt{\dot{x}^2}}{2\lambda} - m\right) = \frac{\eta\sqrt{\dot{x}^2}}{2\lambda}\frac{f}{\lambda}\left(\frac{\dot{x}\cdot\ddot{x}}{\dot{x}^2} - \frac{\dot{\lambda}}{\lambda}\right)c \doteq 0,$$

$$\delta_s\frac{p\cdot x}{\eta\sqrt{p^2}} = \delta_s\frac{\eta\dot{x}\cdot x}{\sqrt{\dot{x}^2}} = \eta\frac{\dot{x}\cdot x}{\sqrt{\dot{x}^2}}\frac{f}{\lambda}\left(\frac{\dot{x}^2 + x\cdot\ddot{x}}{\dot{x}\cdot x} - \frac{\dot{x}\cdot\ddot{x}}{\dot{x}^2}\right)c \doteq \eta\sqrt{\dot{x}^2}\frac{f}{\lambda}c,$$

$$\delta_s\mathbf{k} = \delta_s\eta\frac{m\dot{\mathbf{x}}}{\sqrt{\dot{x}^2}} = \frac{\eta m}{\sqrt{\dot{x}^2}}\frac{f}{\lambda}\left(\ddot{\mathbf{x}} - \frac{\dot{x}\cdot\ddot{x}}{\dot{x}^2}\dot{\mathbf{x}}\right)c \doteq 0, \tag{18}$$

$$\delta_s\mathbf{z} = \delta_s\frac{\eta\sqrt{\dot{x}^2}}{2m\lambda}\left(\mathbf{x} - \frac{\dot{\mathbf{x}}}{\dot{x}^0}x^0\right) = \frac{\eta\sqrt{\dot{x}^2}}{2m\lambda}\frac{f}{\lambda}\left[\left(\frac{\dot{x}\cdot\ddot{x}}{\dot{x}^2} - \frac{\dot{\lambda}}{\lambda}\right)\left(\mathbf{x} - \frac{\dot{\mathbf{x}}}{\dot{x}^0}x^0\right) - \frac{x^0}{\dot{x}^0}\left(\ddot{\mathbf{x}} - \frac{\dot{\mathbf{x}}}{\dot{x}^0}\ddot{x}^0\right)\right]c \doteq 0,$$

where "$\doteq$" means by using $L_\mu \doteq 0$ of Eqs. (6). Therefore, on the extremals of L_0 the gauge variables are λ and $\eta\dot{x}\cdot x/\sqrt{\dot{x}^2}$.

Let us now define the most general BRS gauge-fixed Lagrangian which can be associated to L if we choose a certain BRS-allowed parametrization f of the gauge transformations:

$$\tilde{L} = L_{f\psi\eta} = L + \delta_s\left[\bar{c}\left(\psi(\dot{x}^2,\lambda,\dot{\lambda}) - \frac{b}{2}\right)\right] + \frac{d}{d\tau}[\bar{c}\eta(\dot{x}^2,\lambda,\dot{\lambda},c,\dot{c})]$$
$$= L - b^2/2 + b\psi + \dot{\bar{c}}\eta + \bar{c}(\dot{\eta} - \delta_s\psi), \tag{19}$$

where ψ is the Lagrangian gauge fixing, $b(\tau)$ is a nonlinear Lagrange multiplier (often referred to as the Stückelberg field), $\bar{c}$ is an odd variable (the antighost), and η is an odd function. In the following we shall use $\tilde{L}$ as a shorthand for $L_{f\psi\eta}$. The BRS transformation properties of $\bar{c}$ and b are

$$\delta_s\bar{c} = b,$$
$$\delta_s b = 0. \tag{20}$$

Since the Euler–Lagrange equation for b is $\psi - b \doteq 0$, b could be eliminated and we could use the Lagrangian:

$$\tilde{L}' = L + \tfrac{1}{2}\psi^2 + \dot{\bar{c}}\eta + \bar{c}(\dot{\eta} - \delta_s\psi). \tag{21}$$

But now we should use $\delta_s\bar{c} = \psi$ with $\delta_s^2\bar{c} = \delta_s\psi = -\tilde{L}'_{\bar{c}} \doteq 0$, where $\tilde{L}'_{\bar{c}} = \tilde{L}_{\bar{c}} \doteq 0$ is the Euler–Lagrange equation for $\bar{c}$ and the nilpotency of δ_s would hold only on the extremals.[6] As said at the end of the previous section, one should remember the existence of the two sectors $b \neq 0$ and $b = 0$. Let us remark that if in Eq. (19) we drop the term $-b^2/2$, then we come upon the case $L' = L + \alpha(\tau)\psi$ discussed at the end of Sec. I.

Equation (21) shows that ψ is just the Lagrangian gauge fixing of Sec. I. Instead, Eq. (19) shows that the antighost $\bar{c}$ is a Lagrange multiplier if we choose $\eta = 0$, but becomes a dynamical variable if $\eta \neq 0$: its associated momentum is $\bar{\mathcal{P}} = \eta$.

The form of the Lagrangian (19) and Eqs. (20) imply

$$\delta_s\tilde{L} = \delta_s x^\mu\tilde{L}_\mu + \delta_s\lambda\tilde{L}_\lambda + \delta_s c\tilde{L}_c + \delta_s\bar{c}\tilde{L}_{\bar{c}} + \frac{d}{d\tau}\left(\delta_s x^\mu\frac{\partial\tilde{L}}{\partial\dot{x}^\mu} + \delta_s\lambda\frac{\partial\tilde{L}}{\partial\dot{\lambda}} + \delta_s c\frac{\partial\tilde{L}}{\partial\dot{c}} + \delta_s\bar{c}\frac{\partial\tilde{L}}{\partial\dot{\bar{c}}}\right)$$
$$\equiv \frac{d}{d\tau}\left[f\frac{L}{\lambda}c + \delta_s(\bar{c}\eta)\right]$$
$$= \frac{d}{d\tau}\left[f\frac{L}{\lambda}c + b\eta + \bar{c}\delta_s\eta\right], \tag{22}$$

where $\tilde{L}_\mu,\ldots,$ are the Euler–Lagrange equations for $\tilde{L}$.

The first Nœther theorem produces as a constant of the motion the BRS charge

$$\Omega = \delta_s x^\mu\frac{\partial\tilde{L}}{\partial\dot{x}^\mu} + \delta_s\lambda\frac{\partial\tilde{L}}{\partial\dot{\lambda}} + \delta_s c\frac{\partial\tilde{L}}{\partial\dot{c}}$$

$$+\delta_s\bar{c}\frac{\partial\tilde{L}}{\partial\dot{c}}-f\frac{L}{\lambda}c-b\eta-\bar{c}\delta_s\eta,\quad \dot{\Omega}\doteq 0. \tag{23}$$

Let us remark that in phase space $\overline{\Omega}=\Omega$ is the generator of the Nœther BRS transformations (13), (15), (20): since $\delta\overline{A}=\{\overline{A},\rho\overline{\Omega}\}=\rho\delta_s\overline{A}$, we get $\delta_s\overline{A}=\{\overline{A},\overline{\Omega}\}$ for $\overline{A}$ even, $\delta_s\overline{A}=-\{\overline{A},\overline{\Omega}\}$ for $\overline{A}$ odd.

However, for every given allowed f we have to put restrictions on the functions ψ and η: (i) they must be local functions of their arguments; (ii) $\tilde{L}$ must be independent from the accelerations. We have to study the term $\dot{\eta}-\delta_s\psi$ and restrict it to be a local function of $\dot{x}^2,\lambda,\dot{\lambda},c,\dot{c}$. We get

$$\begin{aligned}\dot{\eta}-\delta_s\psi=&\,2\dot{x}\cdot\ddot{x}\frac{\partial\eta}{\partial\dot{x}^2}+\dot{\lambda}\frac{\partial\eta}{\partial\lambda}+\ddot{\lambda}\frac{\partial\eta}{\partial\dot{\lambda}}+\dot{c}\frac{\partial\eta}{\partial c}+\ddot{c}\frac{\partial\eta}{\partial\dot{c}}\\&-2\frac{\partial\psi}{\partial\dot{x}^2}\dot{x}_\mu\left[\frac{1}{\lambda}\left(f\ddot{x}^\mu+\left(\dot{f}-\frac{\dot{\lambda}}{\lambda}f\right)\dot{x}^\mu\right)c\right.\\&\left.+f\frac{\dot{x}^\mu}{\lambda}\dot{c}\right]-\frac{\partial\psi}{\partial\lambda}(\dot{f}c+f\dot{c})\\&-\frac{\partial\psi}{\partial\dot{\lambda}}(\ddot{f}c+2\dot{f}\dot{c}+f\ddot{c}).\end{aligned}\tag{24}$$

With $f=f(\dot{x}^2,\lambda)$ we get four conditions (of which the first is the coefficient of $\ddot{c}$, the second of $\ddot{\lambda}$, and the third of $\dot{x}\cdot\ddot{x}$):

$$\begin{aligned}&\frac{\partial\eta}{\partial\dot{c}}-f\frac{\partial\psi}{\partial\dot{\lambda}}=0,\quad \frac{\partial\eta}{\partial\dot{\lambda}}-\frac{\partial f}{\partial\lambda}\frac{\partial\psi}{\partial\dot{\lambda}}c=0,\\&\frac{\partial\eta}{\partial\dot{x}^2}-\left[\frac{1}{\lambda}\left(f+2\dot{x}^2\frac{\partial f}{\partial\dot{x}^2}\right)\frac{\partial\psi}{\partial\dot{x}^2}\right.\\&\qquad\left.+\frac{\partial f}{\partial\dot{x}^2}\frac{\partial\psi}{\partial\lambda}+2\dot{\lambda}\frac{\partial^2 f}{\partial\lambda\,\partial\dot{x}^2}\frac{\partial\psi}{\partial\dot{\lambda}}\right]c-2\frac{\partial f}{\partial\dot{x}^2}\frac{\partial\psi}{\partial\dot{\lambda}}\dot{c}=0,\\&\left[\frac{\partial f}{\partial\dot{x}^2}(\ddot{x}^2+\dot{x}\cdot\dddot{x})+2(\dot{x}\cdot\ddot{x})^2\frac{\partial^2 f}{\partial(\dot{x}^2)^2}\right]\frac{\partial\psi}{\partial\dot{\lambda}}c=0.\end{aligned}\tag{25}$$

The solution of the first of Eqs. (25) is $\eta=A(\dot{x}^2,\lambda,\dot{\lambda})c+f(\partial\psi/\partial\dot{\lambda})\dot{c}$. The second of Eqs. (25) then implies the following two conditions:

$$\begin{aligned}&f\frac{\partial^2\psi}{\partial\dot{\lambda}^2}=0\\&\qquad\Rightarrow\psi=\psi_0(\dot{x}^2,\lambda)+\dot{\lambda}\psi_1(\dot{x}^2,\lambda)\Rightarrow\eta=Ac+f\psi_1\dot{c},\\&\frac{\partial A}{\partial\dot{\lambda}}-\frac{\partial f}{\partial\lambda}\frac{\partial\psi}{\partial\dot{\lambda}}=0\Rightarrow A=B(\dot{x}^2,\lambda)+\dot{\lambda}\frac{\partial f}{\partial\lambda}\psi_1(\dot{x}^2,\lambda).\end{aligned}\tag{26}$$

The remaining Eqs. (25) become

$$\begin{aligned}&\frac{\partial B}{\partial\dot{x}^2}-\frac{1}{\lambda}\left(f+2\dot{x}^2\frac{\partial f}{\partial\dot{x}^2}\right)\frac{\partial\psi_0}{\partial\dot{x}^2}-\frac{\partial f}{\partial\dot{x}^2}\frac{\partial\psi_0}{\partial\lambda}=0,\\&\frac{\partial^2 f}{\partial\lambda\,\partial\dot{x}^2}\psi_1-\frac{\partial f}{\partial\lambda}\frac{\partial\psi_1}{\partial\dot{x}^2}+\frac{\partial f}{\partial\dot{x}^2}\frac{\partial\psi_1}{\partial\lambda}\\&\qquad+\frac{1}{\lambda}\left(f+2\dot{x}^2\frac{\partial f}{\partial\dot{x}^2}\right)\frac{\partial\psi_1}{\partial\dot{x}^2}=0,\\&f\frac{\partial\psi_1}{\partial\dot{x}^2}-\frac{\partial f}{\partial\dot{x}^2}\psi_1=0,\\&\text{either}\quad\frac{\partial f}{\partial\dot{x}^2}=0\quad\text{or}\quad\psi_1=0,\end{aligned}\tag{27}$$

with the following two sets of solutions:

$$\begin{aligned}&(1)\quad\frac{\partial f(\dot{x}^2,\lambda)}{\partial\dot{x}^2}\neq0\Rightarrow\delta_s c=\left(\frac{2\dot{x}^2}{\lambda}\frac{\partial f}{\partial\dot{x}^2}+\frac{\partial f}{\partial\lambda}\right)c\dot{c},\\&\psi=\psi_0(\dot{x}^2,\lambda),\\&\eta=B(\dot{x}^2,\lambda)c,\quad\text{with}\\&\frac{\partial B}{\partial\dot{x}^2}=\frac{\partial f}{\partial\dot{x}^2}\frac{\partial\psi_0}{\partial\lambda}+\frac{1}{\lambda}\left(f+2\dot{x}^2\frac{\partial f}{\partial\dot{x}^2}\right)\frac{\partial\psi_0}{\partial\dot{x}^2},\\&\dot{\eta}-\delta_s\psi=\dot{\lambda}\left(\frac{\partial B}{\partial\lambda}-\frac{2\dot{x}^2}{\lambda}\left(\frac{\partial f}{\partial\lambda}-\frac{f}{\lambda}\right)\frac{\partial\psi_0}{\partial\dot{x}^2}-\frac{\partial f}{\partial\lambda}\frac{\partial\psi_0}{\partial\lambda}\right)c\\&\qquad+\left[B-f\left(\frac{\partial\psi_0}{\partial\lambda}+\frac{2\dot{x}^2}{\lambda}\frac{\partial\psi_0}{\partial\dot{x}^2}\right)\right]\dot{c},\\&\delta_s\eta=\left[\frac{2\dot{x}^2}{\lambda}\left(\frac{\partial f}{\partial\dot{x}^2}B-f\frac{\partial B}{\partial\dot{x}^2}\right)+\frac{\partial f}{\partial\lambda}B-f\frac{\partial B}{\partial\lambda}\right]c\dot{c},\end{aligned}\tag{28}$$

$$\begin{aligned}&(2)\quad f=f(\lambda)\Rightarrow\delta_s c=\frac{df(\lambda)}{d\lambda}c\dot{c},\\&\psi=\psi_0(\dot{x}^2,\lambda)+\dot{\lambda}\psi_1(\lambda),\\&\eta=\left[B(\dot{x}^2,\lambda)+\dot{\lambda}\frac{df(\lambda)}{d\lambda}\psi_1(\lambda)\right]c+f(\lambda)\psi_1(\lambda)\dot{c},\\&B=a(\lambda)+\frac{f(\lambda)}{\lambda}\psi_0(\dot{x}^2,\lambda),\\&\dot{\eta}-\delta_s\psi=\dot{\lambda}\left[\frac{\partial B}{\partial\lambda}-\frac{2\dot{x}^2}{\lambda}\left(\frac{df}{d\lambda}-\frac{f}{\lambda}\right)\frac{\partial\psi_0}{\partial\dot{x}^2}-\frac{df}{d\lambda}\frac{\partial\psi_0}{\partial\lambda}\right]c\\&\qquad+\left[B-f\left(\frac{\partial\psi_0}{\partial\lambda}+\frac{2\dot{x}^2}{\lambda}\frac{\partial\psi_0}{\partial\dot{x}^2}\right)\right]\dot{c},\\&\delta_s\eta=\left[\frac{df}{d\lambda}a-f\frac{da}{d\lambda}-\frac{f^2}{\lambda}\left(\frac{\partial\psi_0}{\partial\lambda}-\frac{\psi_0}{\lambda}\right)\right]c\dot{c},\end{aligned}\tag{29}$$

where in Eqs. (28) B is a solution, local in the variables λ and $\dot{x}^2$, of the included differential equation and, in Eqs. (29), $a(\lambda)$ is an arbitrary function. If instead we insist on having $\eta=0$ in Eqs. (25)–(27), i.e., that $\bar{c}$ is a Lagrange multiplier, then the previous solutions have to be restricted as follows [the restriction comes from the $\dot{x}\cdot\ddot{x}$ dependence of $\delta_s\psi$ induced from $\delta_s\dot{x}^2$, due to Eqs. (13)]:

$$\begin{aligned}&(1_0)\quad\frac{\partial f(\dot{x}^2,\lambda)}{\partial\dot{x}^2}\neq0,\\&\psi=\psi_0(\dot{x}^2,\lambda)\quad\text{solution of}\quad\frac{\partial f}{\partial\dot{x}^2}\frac{\partial\psi_0}{\partial\lambda}\\&\qquad+\frac{1}{\lambda}\left(f+2\dot{x}^2\frac{\partial f}{\partial\dot{x}^2}\right)\frac{\partial\psi_0}{\partial\dot{x}^2}=0,\\&\delta_s\psi=\dot{\lambda}\left[\frac{2\dot{x}^2}{\lambda}\left(\frac{\partial f}{\partial\lambda}-\frac{f}{\lambda}\right)\frac{\partial\psi_0}{\partial\dot{x}^2}+\frac{\partial f}{\partial\lambda}\frac{\partial\psi_0}{\partial\lambda}\right]c\\&\qquad+f\left(\frac{2\dot{x}^2}{\lambda}\frac{\partial\psi_0}{\partial\dot{x}^2}+\frac{\partial\psi_0}{\partial\lambda}\right)\dot{c},\end{aligned}\tag{30}$$

$$\begin{aligned}&(2_0)\quad f=f(\lambda),\\&\psi=\psi_0(\lambda),\\&\delta_s\psi=\dot{\lambda}\frac{df}{d\lambda}\frac{d\psi_0}{d\lambda}c+f\frac{d\psi_0}{d\lambda}\dot{c}.\end{aligned}\tag{31}$$

In Eqs. (28) and (30), $\psi_0(\dot{x}^2,\lambda)$ must be a solution, again local in its arguments, of the corresponding differential equation; (1_0) is a special case of (1) when $B=0$ and ψ_0 is one of the said solutions, while (2_0) is a special case of (2) when $B=\psi_1=0$ and $\partial\psi_0/\partial\dot{x}^2=0$. The result we have obtained is that, once f is given, the Lagrangian gauge fixing ψ and the function η are severely restricted by the need to have a Lagrangian independent from the accelerations. Therefore, even when $\eta = 0$ we cannot add an arbitrary Lagrangian gauge fixing ψ.

Let us remark that in between (1) and (1_0) there is the case

$$(1') \quad \frac{\partial f(\dot{x}^2,\lambda)}{\partial \dot{x}^2} \neq 0,$$

$$\psi = \psi_0(\dot{x}^2,\lambda)$$

$$\text{solution of} \quad \frac{\partial f}{\partial \dot{x}^2}\frac{\partial \psi_0}{\partial \lambda} + \frac{1}{\lambda}\left(f + 2\dot{x}^2 \frac{\partial f}{\partial \dot{x}^2}\right)\frac{\partial \psi_0}{\partial \dot{x}^2} = 0,$$

$$\eta = B(\lambda)c, \quad B(\lambda)\neq 0 \quad \text{arbitrary},$$

$$\dot{\eta} - \delta_s \psi = \dot{\lambda}\left[\frac{dB}{d\lambda} - \frac{2\dot{x}^2}{\lambda}\left(\frac{\partial f}{\partial \lambda} - \frac{f}{\lambda}\right)\frac{\partial \psi_0}{\partial \dot{x}^2} - -\frac{\partial f}{\partial \lambda}\frac{\partial \psi_0}{\partial \lambda}\right] c + \left[B - f\left(\frac{\partial \psi_0}{\partial \lambda} + \frac{2\dot{x}^2}{\lambda}\frac{\partial \psi_0}{\partial \dot{x}^2}\right)\right]\dot{c},$$

$$\delta_s \eta = \left[\frac{2\dot{x}^2}{\lambda}\frac{\partial f}{\partial \dot{x}^2} B + \frac{\partial f}{\partial \lambda} B - f\frac{dB}{d\lambda}\right] c\dot{c}, \tag{32}$$

and in between (2) and (2_0) there is the case

$$(2') \quad f = f(\lambda),$$

$$\psi = \psi_0(\lambda)$$

$$\eta = B(\lambda)c, \quad B(\lambda) = a(\lambda) + \frac{f(\lambda)}{\lambda}\psi_0(\lambda),$$

$$\dot{\eta} - \delta_s \psi = \dot{\lambda}\left[\frac{dB(\lambda)}{d\lambda} - \frac{df}{d\lambda}\frac{d\psi_0}{d\lambda}\right] c + \left(B - f\frac{d\psi_0}{d\lambda}\right)\dot{c},$$

$$\delta_s \eta = \left[\frac{df}{d\lambda} a - f\frac{da}{d\lambda} - \frac{f^2}{\lambda}\left(\frac{d\psi_0}{d\lambda} - \frac{\psi_0}{\lambda}\right)\right] c\dot{c}. \tag{33}$$

Let us remark that $\widetilde{L} = L - b^2/2 + b\psi - \bar{c}\delta_s\psi$ for (1_0), (2_0) and $\widetilde{L} = L - b^2/2 + b\psi - \bar{c}\delta_s\psi + (d/d\tau)(\bar{c}\eta)$ for ($1'$), ($2'$) have the same Euler–Lagrange equations, because they have the same ψ and differ for a total τ derivative. But if we call (1_R), (2_R) the classes (1), (2) with ($1'$), ($2'$) excluded, respectively, then there is no analog for them in (1_0), (2_0); (1_R) is the subcase of (1) defined by $\partial B/\partial \dot{x}^2 \neq 0$, while ($2_R$) is the subcase of (2) defined by $\partial\psi_0/\partial\dot{x}^2$, ψ_1, $B = a + f\psi_0/\lambda$ not vanishing simultaneously. (1_R), (2_R) allow us to use more general Lagrangian gauge fixings, which necessarily require the function η, because, without $(d/d\tau)(\bar{c}\eta)$, $\bar{c}\delta_s\psi$ would introduce a dependence on the accelerations.

For the sake of simplicity we shall restrict ourselves from now on to the gauge transformations parametrized by $f = k + h\lambda$ ($h = 0$ is the parametrization usually chosen, see for instance Ref. 7), for which $\delta_s c = hc\dot{c}$. Then (2_R) and (2_0) reduce to

$$(A_R) \quad \psi = \psi_0(\dot{x}^2,\lambda) + \dot{\lambda}\psi_1(\lambda),$$

$$\eta = [a(\lambda) + h\dot{\lambda}\psi_1(\lambda) + [(k + h\lambda)/\lambda]\psi_0(\dot{x}^2,\lambda)]c + (k + h\lambda)\psi_1(\lambda)\dot{c}, \tag{34}$$

$$(A_0) \quad \psi = \chi(\lambda).$$

To simplify the matter we shall also restrict (A_R) to $\psi_0 = \chi(\lambda)$, $B = a + [(k + h\lambda)/\lambda]\psi_0 = 0$, $\psi_1 = \alpha$, so that we shall use

$$(A'_R) \quad \psi = \chi(\lambda) + \alpha\dot{\lambda}, \quad \alpha \neq 0,$$

$$\eta = \alpha[(k + h\lambda)\dot{c} + h\dot{\lambda}c],$$

$$\widetilde{L}_R = L - b^2/2 + b(\chi + \alpha\dot{\lambda}) + \alpha(k + h\lambda)\dot{\bar{c}}\dot{c} + \alpha h\dot{\lambda}\dot{\bar{c}}c - h\dot{\lambda}\chi'\bar{c}c - (k + h\lambda)\chi'\bar{c}\dot{c}$$

$$= \widetilde{L}_0 + \alpha[b\dot{\lambda} + (k + h\lambda)\dot{\bar{c}}\dot{c} + h\dot{\lambda}\dot{\bar{c}}c], \tag{35}$$

$$(A'_0) \quad \psi = \chi(\lambda),$$

$$\widetilde{L}_0 = L - b^2/2 + b\chi - \chi'[h\dot{\lambda}\bar{c}c + (k + h\lambda)\bar{c}\dot{c}], \tag{36}$$

where if $A = A(\lambda)$ then $A'(\lambda) = dA(\lambda)/d\lambda$. In Eqs. (36) we must require $\chi' \neq 0$. Let us now study the Hamiltonian description associated with the inequivalent Lagrangians $\widetilde{L}_0$ and $\widetilde{L}_R$, corresponding to the two different ways of introducing Lagrangian BRS gauge fixings of L when the gauge transformations are parametrized with $f = k + h\lambda$. As we shall see, a Kato–Ogawa[8] like Hamiltonian theory will correspond to $\widetilde{L}_0$, while $\widetilde{L}_R$ will produce the BFV[9,10] approach.

Let us start with $\widetilde{L}_0$, whose momenta, canonical Hamiltonian and Euler–Lagrange equations have the following expressions:

$$p_{0\mu} = \frac{\partial \widetilde{L}_0}{\partial \dot{x}^\mu} = p_\mu = \frac{\dot{x}_\mu}{2\lambda},$$

$$\pi_{0\lambda} = \frac{\partial \widetilde{L}_0}{\partial \dot{\lambda}} = -h\chi'\bar{c}c,$$

$$\pi_b = \frac{\partial \widetilde{L}_0}{\partial \dot{b}} = 0, \quad \Rightarrow \overline{H}_0 = \lambda(p^2 - m^2) + \frac{b^2}{2} - b\chi,$$

$$\mathscr{P}_0 = \frac{\partial \widetilde{L}_0}{\partial \dot{c}} = (k + h\lambda)\chi'\bar{c},$$

$$\overline{\mathscr{P}}_0 = \frac{\partial \widetilde{L}_0}{\partial \dot{\bar{c}}} = 0, \tag{37}$$

$$L_{0\mu} = L_\mu = -\frac{d}{d\tau}\frac{\dot{x}_\mu}{2\lambda} \doteq 0,$$

$$L_{0\lambda} = L_\lambda + \chi'(b + h\dot{\bar{c}}c) - (k + h\lambda)\chi''\bar{c}\dot{c}$$

$$= m^2 - \frac{\dot{x}^2}{2\lambda} + \chi'(b + h\dot{\bar{c}}c) - (k + h\lambda)\chi''\bar{c}\dot{c} \doteq 0,$$

$$L_{0b} = \chi - b \doteq 0,$$

$$L_{0c} = -(k + h\lambda)\frac{d}{d\tau}(\chi'\bar{c}) \doteq 0,$$

$$L_{0\bar{c}} = -\chi'[h\dot{\lambda}c + (k + h\lambda)\dot{c}] \doteq 0. \tag{38}$$

The conserved BRS charge is

$$\Omega_0 = (k + h\lambda)(\dot{x}^2/4\lambda^2 - m^2)c$$

$$= (k + h\lambda)(p^2 - m^2)c = \overline{\Omega}_0 \quad \{\overline{\Omega}_0, \overline{\Omega}_0\} = 0. \tag{39}$$

There are four primary constraints:

$$\chi_{0\lambda} = \pi_{0\lambda} + h\chi'\bar{c}c \approx 0, \quad \pi_b \approx 0,$$

$$\rho_0 = \mathscr{P}_0 - (k + h\lambda)\chi'\bar{c} \approx 0, \quad \overline{\mathscr{P}}_0 \approx 0, \tag{40}$$

and the Dirac Hamiltonian is $\overline{H}_D = \overline{H}_0 + \mu_1(\tau)\chi_{0\lambda} + \mu_2(\tau)\pi_b + \mu_3(\tau)\rho_0 + \mu_4(\tau)\overline{\mathscr{P}}_0$, with μ_3 and μ_4 odd Dirac multipliers. The Dirac–Bergmann algorithm for ρ_0 and $\overline{\mathscr{P}}_0$ determines μ_3 and μ_4, while for π_b generates the secondary constraint $b-\chi(\lambda)\approx 0$, which in turn determines μ_2. By using the Poisson brackets $\{c,\mathscr{P}_0\} = \{\bar{c},\overline{\mathscr{P}}_0\} = -1$ we obtain

$$\mu_2 = \chi'\mu_1, \quad \mu_3 = \frac{-h\mu_1}{k+h\lambda}c, \quad \mu_4 = -\frac{\chi''}{\chi'}\mu_1\bar{c}. \tag{41}$$

Then $\chi_{0\lambda}$ generates the secondary constraint $b\chi' - (p^2 - m^2) + h(h\chi' + k\chi'')/(k+h\lambda)\mu_1\bar{c}c \approx 0$, which unusually depends on $\mu_1(\tau)$, and its τ-constancy produces $\mu_1 = 0$, so that $\mu_2 = \mu_3 = \mu_4 = 0$.

In the end we get the following three pairs of 2nd-class constraints (except for the value $b=0$):

$$\begin{aligned} &b-\chi(\lambda)\approx 0, \quad \pi_b\approx 0,\\ &\rho_0 = \mathscr{P}_0 - (k+h\lambda)\chi'(\lambda)\bar{c}\approx 0, \quad \overline{\mathscr{P}}_0\approx 0,\\ &\chi(\lambda)\chi'(\lambda) - (p^2-m^2)\approx 0, \quad \chi_{0\lambda} = \pi_{0\lambda} + h\chi'(\lambda)\bar{c}c\approx 0, \end{aligned} \tag{42}$$

which allow the elimination of the variables b, π_b, $\bar{c}$, $\overline{\mathscr{P}}_0$, λ, $\pi_{0\lambda}$ by means of Dirac brackets. We remain with the variables x^μ, p_μ, c, $\mathscr{P}_0$. This is the situation found by Kato–Ogawa[8] in their BRS approach to the string. However, there is a second sector in which $b-\chi(\lambda)\approx 0$ is satisfied by $b\approx 0$, $\chi(\lambda)\approx 0$ (proliferation of constraints[2]: here, instead of Eqs. (42), we get three pairs of 2nd-class constraints ($b\approx 0$, $\pi_b\approx 0, \rho_0\approx 0, \overline{\mathscr{P}}_0\approx 0, \chi(\lambda)\approx 0, \chi_{0\lambda}\approx 0$) and one 1st-class one $p^2 - m^2\approx 0$ with $\overline{H}'^F_D = \lambda(p^2-m^2)$. This sector reproduces the original theory with λ gauge fixed by $\chi(\lambda)\approx 0$. If $\chi(\lambda)\approx 0$ has d solutions $\lambda(\tau) = \lambda_1,\dots,\lambda_d$ ($\dot\lambda = 0$), this is the residual discrete gauge freedom left by the Lagrangian gauge fixing. By going to Dirac brackets one has $\overline{H}'^F_{Da} = \lambda_a(p^2-m^2)$, $a = 1,\dots,d$ [the Dirac multiplier of the original theory assumes d discrete values $\mu(\tau) = \lambda_a$, $a=1,\dots,d$] and $\overline{\Omega}_{0a} = (k+h\lambda_a)(p^2-m^2)c\approx 0$. This sector corresponds to selecting those special extremals of the gauge-fixed theory which are the subset of the original extremals compatible with the residual gauge freedom. Let us note that this sector is not present when in the Kato–Ogawa approach the Lagrangian gauge fixing ψ is allowed to be velocity dependent [in this model, case (2_0), it is not possible]. In the sector $b\neq 0$ we can rewrite the final Dirac Hamiltonian $\overline{H}^F_D = \overline{H}_0$ in the following form by taking into account Eqs. (42) and (39):

$$\begin{aligned} \overline{H}_0 &= \lambda(p^2-m^2) + \frac{b^2}{2} - b\chi\\ &= \left(\lambda - \frac{p^2-m^2}{2\chi'^2}\right)(p^2-m^2) - \frac{p^2-m^2}{\chi'}\left(\chi - \frac{p^2-m^2}{\chi'}\right)\\ &\quad + \frac{1}{2}\left[(b-\chi)^2 - \left(\chi - \frac{p^2-m^2}{\chi'}\right)^2\right]\\ &= \{\bar\rho_{0\psi},\overline{\Omega}_0\} - \frac{p^2-m^2}{\chi'}\left(\chi - \frac{p^2-m^2}{\chi'}\right)\approx\{\bar\rho_{0\psi},\overline{\Omega}_0\},\\ \bar\rho_{0\psi} &= -\frac{1}{k+h\lambda}\left(\lambda - \frac{p^2-m^2}{2\chi'^2}\right)\mathscr{P}_0. \end{aligned} \tag{43}$$

Since $\{\overline{\Omega}_0,\overline{\Omega}_0\} = 0$, we have obtained a nilpotent BRS charge, a BRS invariant Hamiltonian (modulo the 2nd-class constraints), and the Hamiltonian form $\bar\rho_{0\psi}$ of the Lagrangian gauge fixing $\psi = \chi(\lambda)$. In the sector with $b=0$ we get $\overline{H}'^F_{Da} = \{-\lambda_a/(k+h\lambda),\overline{\Omega}_{0a}\}^*$ and $\{\overline{\Omega}_{0a},\overline{\Omega}_{0a}\}^* = 0$: this is the sector compatible with $p^2 - m^2\approx 0$ in the Kato–Ogawa approach.

If we consider $\tilde{L}_R$, the momenta, the canonical Hamiltonian, and the Euler–Lagrange equations are

$$\left.\begin{aligned} &p_{R\mu} = \frac{\partial\tilde{L}_R}{\partial\dot{x}^\mu} = p_\mu = \frac{\dot{x}_\mu}{2\lambda},\\ &\pi_{R\lambda} = \alpha(b + h\dot{\bar{c}}c) - h\chi'\bar{c}c,\\ &\pi_b = 0\\ &\mathscr{P}_R = (k+h\lambda)(\chi'\bar{c} - \alpha\dot{\bar{c}}),\\ &\overline{\mathscr{P}}_R = \alpha[h\dot\lambda c + (k+h\lambda)\dot{c}], \end{aligned}\right\} \Rightarrow \overline{H}_R = \lambda(p^2-m^2) + \frac{b^2}{2} - b\chi + \frac{\chi'}{\alpha}\bar{c}\overline{\mathscr{P}}_R - \frac{\mathscr{P}_R\overline{\mathscr{P}}_R}{\alpha(k+h\lambda)}, \tag{44}$$

$$\begin{aligned} &L_{R\mu} = L_\mu \doteq 0,\\ &L_{R\lambda} = L_\lambda + b\chi' + \alpha h\dot{\bar{c}}\dot{c} - h\chi''\dot\lambda\bar{c}c - [h\chi' + (k+h\lambda)\chi'']\dot{\bar{c}}c - \frac{d}{d\tau}[\alpha(b + h\dot{\bar{c}}c) - h\chi'\bar{c}c]\doteq 0,\\ &L_{Rb} = \chi + \alpha\dot\lambda - b\doteq 0,\\ &L_{Rc} = h\dot\lambda(\chi'\bar{c} - \alpha\dot{\bar{c}}) - \frac{d}{d\tau}[(k+h\lambda)(\chi'\bar{c} - \alpha\dot{\bar{c}})]\doteq 0,\\ &L_{R\bar{c}} = -\chi'[h\dot\lambda c + (k+h\lambda)\dot{c}] - \frac{d}{d\tau}[\alpha(h\dot\lambda c + (k+h\lambda)\dot{c})]\doteq 0. \end{aligned} \tag{45}$$

The conserved BRS charge is

$$\Omega_R = (k+h\lambda)((\dot{x}^2/4\lambda^2) - m^2)c + \alpha\dot{b}[h\dot\lambda c + (k+h\lambda)\dot{c}] = (k+h\lambda)(p^2-m^2)c + \dot{b}\overline{\mathscr{P}}_R = \overline{\Omega}_R \quad \{\overline{\Omega}_R,\overline{\Omega}_R\} = 0. \tag{46}$$

We now have only two primary 2nd-class constraints:

$$\chi_{R\lambda} = \pi_{R\lambda} - \alpha b + [h/(k+h\lambda)]\mathscr{P}_R c \approx 0,$$
$$\pi_b \approx 0, \tag{47}$$

which imply the elimination of b and π_b and the determination of μ_1, μ_2 in $\overline{H}^F_D = \overline{H}_R + \mu_1\chi_{R\lambda} + \mu_2\pi_b$. If we use Dirac brackets we get

$$\overline{H}_R = \lambda(p^2-m^2) + \frac{1}{2\alpha^2}\left(\pi_{R\lambda} + \frac{h\mathscr{P}_R c}{k+h\lambda}\right)^2 - \frac{\chi}{\alpha}\left(\pi_{R\lambda} + \frac{h\mathscr{P}_R c}{k+h\lambda}\right) + (\chi'/\alpha)\bar{c}\overline{\mathscr{P}}_R - \mathscr{P}_R\overline{\mathscr{P}}_R/\alpha(k+h\lambda),$$
$$\overline{\Omega}_R = (k+h\lambda)(p^2-m^2)c + (1/\alpha)\,(\pi_{R\lambda} + h\mathscr{P}_R c/(k+h\lambda))\overline{\mathscr{P}}_R, \tag{48}$$

and

$$\overline{H}_R = \{\bar{\rho}_{R\psi\eta}, \overline{\Omega}_R\}^*$$
$$\bar{\rho}_{R\psi\eta} = \frac{-\lambda}{k+h\lambda}\mathscr{P}_R + \left[\chi - \frac{1}{2\alpha}\left(\pi_{R\lambda} + \frac{h\mathscr{P}_R c}{(k+h\lambda)}\right)\right]\bar{c}. \tag{49}$$

Again, $\overline{H}_R$ is BRS invariant, $\bar{\rho}_{R\psi\eta}$ is the Hamiltonian version of $\psi = \chi(\lambda) + \alpha\dot{\lambda}$, and $\eta = \alpha[(k+h\lambda)\dot{c} + h\dot{\lambda}c]$. In this case we have recovered the BFV[9]–Henneaux[10] approach. Now, since $\chi + \alpha\dot{\lambda} \doteq 0$, implied by $b=0$, is not a Hamiltonian constraint, the two sectors $b=0$ and $b\neq 0$ are described by the same constraints and the difference between the two appears only at the level of the solutions of the Hamilton equations.

Let us make some general comments. In both cases (35) and (36) with $f(\lambda) = k + h\lambda$ Eqs. (37) and (44) would imply the notations $\mathscr{P}_0^{f\chi}$, $\overline{\mathscr{P}}_0^{f\chi} = 0$ and $\mathscr{P}_R^{f\psi\eta}$, $\overline{\mathscr{P}}_R^{f\psi\eta}$, respectively, because their definition depends on these quantities. Moreover, in both cases we have Euler–Lagrange equations for $c(\tau)$ and $\bar{c}(\tau)$. On the extremals we get $c(\tau) \doteq c_0^{f\chi}(\tau)$, $\bar{c}(\tau) \doteq \bar{c}_0^{f\chi}(\tau)$ and $c(\tau) \doteq c_R^{f\psi\eta}(\tau)$, $\bar{c}(\tau) \doteq \bar{c}_R^{f\psi\eta}(\tau)$, respectively. Therefore, $\rho c_0^{f\chi}(\tau)$, $\rho c_R^{f\psi\eta}(\tau)$ are the BRS analogs of the $\tilde{\epsilon}_\psi(\tau)$ associated with the residual gauge freedom considered in Sec. I. They are a refinement of the $\tilde{\epsilon}_\psi(\tau)$, because now the condition $\delta_s^2 = 0$ (i.e., $\{\overline{\Omega}, \overline{\Omega}\} = 0$) ensures that the residual gauge transformations satisfy the gauge algebra hypothesis. Since $\delta_s\psi = -\tilde{L}_{\bar{c}} \doteq 0$ (the Euler–Lagrangian equation for $\bar{c}$) turns out to be the equation of motion for c [see Eqs. (38) and (45)], we find again as in Sec. I that $\tilde{\epsilon}_\psi(\tau)$ is determined by $\delta_s\psi = 0$.

Moreover, instead of the original constraints $\pi_\lambda \approx 0$, $p^2 - m^2 \approx 0$ (implied by $\dot{\pi}_\lambda \approx 0$) we have the quantities $\pi_{0\lambda}$, $p^2 - m^2$ or $\pi_{R\lambda}$, $p^2 - m^2$ which do not vanish. The original constraint submanifold of phase space is no more known. Therefore, it is improper to say that in $\overline{\Omega}_0 = (k+h\lambda)(p^2-m^2)c$ and

$$\overline{\Omega}_R = (k+h\lambda)(p^2-m^2)c + (1/\alpha)\,(\pi_{R\lambda} + h\mathscr{P}_R c/(k+h\lambda))\overline{\mathscr{P}}_R$$

we have the original constraints. Even formally this should be true for $\overline{\Omega}_R$ only if $h=0$: we recover the BFV–Henneaux interpretation only with simple allowed parametrizations f of the gauge transformations. With the improper interpretation, the Kato–Ogawa case would correspond to the elimination of the primary constraint $\pi_\lambda \approx 0$ associated to the einbein λ, leaving only the relevant secondary constraint $p^2 - m^2 \approx 0$. The 2nd-class constraints of Eqs. (42) are the equivalent of the addition of a Dirac gauge fixing $\chi(\lambda) \approx 0$ to the original model with L. The Kato–Ogawa form of the BRS charge $\overline{\Omega}_0$ corresponds, in the case of a set $\pi_A \approx 0$, $\phi_A \approx 0$ of 1st class constraints, to considering only the ϕ_A's: this is the form of the BRS charge used in the theoretical discussions.

However, this improper interpretation is valid in the sense that $\overline{\Omega}$ is the generator of the BRS transformations $\delta_s q^i$, generalizing the residual gauge transformations $\tilde{\delta}_0 q^i$, after the addition of the Lagrangian gauge fixing, via $\epsilon(\tau) \mapsto \rho c(\tau)$.

In the original model we had $\pi_\lambda \approx 0$ and $\dot{\pi}_\lambda = \{\pi_\lambda, \overline{H}_D\} = -(p^2-m^2) \approx 0$. If in the improper interpretation of the BFV case we impose $\pi_{R\lambda} + h\mathscr{P}_R\overline{\mathscr{P}}_R/(k+h\lambda) \approx 0$ we have from Eqs. (48)

$$\{\pi_{R\lambda} + h\mathscr{P}_R\overline{\mathscr{P}}_R/(k+h\lambda), \overline{H}_R\}^*$$
$$= -(p^2-m^2) + (\chi'/\alpha)\,(\pi_{R\lambda} + h\mathscr{P}_R c/(k+h\lambda))$$
$$-(\chi''/\alpha)\bar{c}\overline{\mathscr{P}}_R \approx -(p^2-m^2) - (\chi''/\alpha)\,\bar{c}\overline{\mathscr{P}}_R.$$

If we ask for its vanishing, we recover $p^2 - m^2 \approx 0$ only if $\chi''(\lambda) = 0$. This is the Govaert's condition[11] for the free-relativistic particle path integral. In its evaluation following the prescriptions of Ref. 10 one should do an integration on the quantity, called modulus, $c_\lambda = \int_{\tau_1}^{\tau_2} d\tau\,\lambda(\tau)$ (the total proper time) with $c_\lambda \in (-\infty, +\infty)$. Therefore, c_λ depends on the Lagrangian gauge fixing $\psi = \alpha\dot{\lambda} + \chi(\lambda) = 0$ (the Teichmuller space of the model, subsequently reduced to the modular space $c_\lambda \in [0, +\infty)$). Actually, instead of an integral over dc, one has to do an integral over $\lambda_N = \lambda(\tau_2)$ deriving from the measure $D\lambda(\tau)$: the solution of $\psi = 0$ now gives a λ_N spanning only once the whole range $(-\infty, +\infty)$ only for $\chi'' = 0$.

We end this section by noting that in the BFV case the Hamiltonian gauge fixing $\bar{\rho}_{R\psi\eta}$ of Eq. (49) has a restricted form. What happens if we choose a different $\bar{\rho}_R$ in Eq. (49) as in the usual BFV approach? Clearly, the connection with $\tilde{L}_R$ of Eq. (35) is lost, because we have a different Hamiltonian $\overline{H}'_R$. We can start from this Hamiltonian, solve the first half of the Hamilton equations to find new functions giving the momenta in terms of the velocities, and look for $\tilde{L}'_R$. Due to the different relation between momenta and velocities, $\overline{\Omega}_R$ will define a new operator δ'_s so that $\delta'_s q^i \neq \delta_s q^i$ of Eqs. (13). This implies that $\tilde{L}'_R$ will have the form $\tilde{L}'_R = L' + \delta'_s$

$[\bar{c}(\psi' - b/2)] + (d/d\tau)(\bar{c}\eta')$ with $L' = L + L_\rho$ such that $\delta'_s L' = dF'/d\tau$. When $\bar{\rho}_R = \bar{\rho}_{R\psi\eta}$, then we recover $\delta'_s = \delta_s$ and $L_\rho = 0$. It is clear that the use of $\bar{\rho}_R \neq \bar{\rho}_{R\psi\eta}$, natural in phase space, is very complicated at the Lagrangian level, because we also have to do a functional change of the form of the original Lagrangian.

III. MORE ON THE FREE PARTICLE

In this section, we shall discuss various other aspects of the two main variants of the BRS approach.

In the BFV case we have, from Eqs. (48),

$$\begin{aligned}
\delta_s \mathscr{P}_R &= -\{\mathscr{P}_R, \bar{\Omega}_R\}^* \\
&= (k + h\lambda)(p^2 - m^2) - \frac{h}{\alpha}\frac{\mathscr{P}_R \bar{\mathscr{P}}_R}{k + h\lambda} \leftrightarrow \frac{p\cdot x}{\eta\sqrt{p^2}}, \\
\delta_s p\cdot x/\eta\sqrt{p^2} &= \{p\cdot x/\eta\sqrt{p^2}, \bar{\Omega}_R\}^* \\
&= 2\eta\sqrt{p^2}(k + h\lambda)c \leftrightarrow \mathscr{P}_R, \\
\delta_s \lambda &= \{\lambda, \bar{\Omega}_R\}^* = (1/\alpha)\bar{\mathscr{P}}_R \leftrightarrow \bar{c}, \\
\delta_s \bar{c} &= -\{\bar{c}, \bar{\Omega}_R\}^* = \frac{1}{\alpha}(\pi_{R\lambda} + h\mathscr{P}_R c/(k + h\lambda)) \leftrightarrow \lambda.
\end{aligned} \tag{50}$$

For $h = 0$, Eqs. (50) imply that $p^2 - m^2$, c, $\bar{\mathscr{P}}_R$, $\pi_{R\lambda}$ are trivial BRS observables [Eqs. (50) gives their generalization for $h \neq 0$]: they have vanishing Poisson brackets with the BRS charge but are irrelevant for the search of the physical observables, which are equivalence classes of BRS observables just modulo the trivial BRS observables. Therefore, their conjugate variables $p\cdot x/\eta\sqrt{p^2}, \mathscr{P}_R, \bar{c}, \lambda$ [shown in the second column of Eqs. (50)] are to be interpreted as gauge variables: indeed the even ones are the gauge variables of the original model. This justifies the terminology according to which c, $\bar{\mathscr{P}}_R$ are negative degrees of freedom erasing $p^2 - m^2$ and π_λ. This is the Kugo–Ojima quartet mechanism,[12] basis of the Parisi–Sourlas approach[13] in field theory.

Let us now consider the anti-BRS transformations[14] in the standard case $f = k$ and with $\tilde{L}_R$ of Eqs. (35):

$$\begin{array}{lll}
\text{BRS} & \text{Anti-BRS} & \\
\delta_s x^\mu = k(\dot{x}^\mu/\lambda)c, & \tilde{\delta}_s x^\mu = k(\dot{x}^\mu/\lambda)\bar{c}, & \\
\delta_s \lambda = k\dot{c}, & \tilde{\delta}_s \lambda = k\dot{\bar{c}}, & \delta_s^2 = \tilde{\delta}_s^2 = 0, \\
\delta_s c = 0, & \tilde{\delta}_s \bar{c} = 0, & \delta_s\tilde{\delta}_s + \tilde{\delta}_s\delta_s = 0. \\
\delta_s \bar{c} = b, & \tilde{\delta}_s c = -b, & \\
\delta_s b = 0, & \tilde{\delta}_s b = 0, &
\end{array} \tag{51}$$

However, we get

$$\tilde{\delta}_s \tilde{L}_R = \frac{d}{d\tau}\left(k\frac{L}{\lambda}\bar{c} + \alpha k b\dot{\bar{c}}\right) + k\chi'(b\dot{\bar{c}} - \dot{b}\bar{c}) + k\chi''\bar{c}\,\dot{\bar{c}}c. \tag{52}$$

Therefore, only for $\chi'(\lambda) = 0$, i.e., $\psi = a + \alpha\dot{\lambda}$ in Eqs. (36), we have a quasiinvariance with the conserved anti-BRS charge $\bar{\bar{\Omega}}_R = k(p^2 - m^2)\bar{c} - (1/\alpha)\pi_{R\lambda}\mathscr{P}_R$. Since $\bar{\Omega}_R = k(p^2 - m^2)c + (1/\alpha)\pi_{R\lambda}\bar{\mathscr{P}}_R$, we get $\{\bar{\bar{\Omega}}_R, \bar{\bar{\Omega}}_R\} = \{\bar{\Omega}_R, \bar{\Omega}_R\} = \{\bar{\bar{\Omega}}_R, \bar{\Omega}_R\} = \{\bar{\Omega}_R, \bar{H}_R\} = \{\bar{\bar{\Omega}}_R \bar{H}_R\} = 0$. In any case, also when $\chi'(\lambda) \neq 0$ so that the previous $\bar{\bar{\Omega}}_R$ is not a constant of the motion, one can algebraically build a linear realization of the group Osp (1,1,2)[7,15–17] (which is not an invariance group of $\tilde{L}_R$): two of its generators are linear combinations of $\bar{\Omega}_R, \bar{\bar{\Omega}}_R$ and the eight variables supporting this linear realization are proportional to c, $\mathscr{P}_R$, $\bar{c}$, $\bar{\mathscr{P}}_R$, λ, $\pi_{R\lambda}$, $p\cdot x/p^2$, $\frac{1}{2}(p^2 - m^2)$. By englobing the Poincaré group in D dimensions, this construction is enlarged to obtain the group IOsp (D,2,2), also containing the Parisi–Sourlas supersymmetry generators connecting the members of the previous quartets. This is the invariance group of a field theory action deduced from the first quantized theory[7] and whose quantization is equivalent to the usual one for the Klein–Gordon field due to the Parisi–Sourlas mechanism.

Let us add some remarks about the path integral quantization.[9,10,18] When one uses a Lagrangian gauge fixing and the BRS Lagrangian with $\eta = 0$ (Kato–Ogawa formalism), the ghost and the antighost are introduced to exponentiate a Jacobian, which roughly speaking is introduced to eliminate the infinite volume of the gauge group and which describes the gauge transformations properties of the Lagrangian gauge fixing. It can be recovered from the phase space path integral with the Faddeev–Popov measure associated with a family of Dirac gauge-fixing constraints, each one depending on an arbitrary function, by making a Gaussian average over these functions (this is the transition from a completely gauge-fixed situation to the approach with the Lagrangian gauge fixings and the associated residual gauge freedom).

In the $\eta \neq 0$ case the BFV approach recovers the Lagrangian path integral starting from the phase space integral: moreover, by doing a certain contraction,[9,10] which is equivalent to completely fixing the residual gauge freedom, it can recover the phase space path integral with Faddeev–Popov measure corresponding to those Dirac gauge-fixing constraints performing the same complete fixation. In this limit, the scaled ghosts and antighosts become the Faddeev–Popov ones needed to exponentiate the determinant of the Poisson brackets of the original constraints with the Dirac gauge fixings.

In the original BRS ($\eta = 0$) and BFV ($\eta \neq 0$) approaches no restriction is put on the boundary conditions for the path integral and the submanifold of the original constraints is not selected in the extended configuration/phase space. The philosophy is that the expectation value of those quantities, which were gauge invariant for the original theory, now depend neither on the Lagrangian gauge fixing ψ nor on its phase space counterpart $\bar{\rho}_{\psi\eta}$ (see, for instance, the Fradkin–Vilkovisky theorem[9,10]). This amounts to a selection of the gauge-invariant quantities as is done in Eqs. (18) for the relativistic particle. If we do a BRS transformation, which is unitarily implementable in the extended phase space, the path integrals before and after such a transformation should be connected in the standard way[19]: the formal Fradkin–Vilkovisky theorem is applicable just in these situations.

There is another possibility. In Sec. I we said that by choosing a special class of initial data for the gauge-fixed Lagrangian we could restrict ourselves to a subclass of the extremals of the original Lagrangian, those which have their gauge freedom reduced to the residual one. In phase space, this amounts to projecting on the constraint submanifold and again to restricting the gauge freedom to the residual one. Therefore, when we add to the BFV path integral a set of BRS invariant boundary conditions as proposed by Henneaux–Teitelboim,[20,10] we are just restricting ourselves to the constraint submanifold and to the residual gauge freedom.

Let us end this section by noting that to the Lagrangians $L_\eta = \eta m\sqrt{\dot{x}^2}, \eta = \pm$, Lagrangian and BRS gauge-fixing terms can be added as shown in Sec. II, but only with $\eta = 0$ because there are no einbeins and the BFV approach cannot be applied. The BRS transformations

$$\begin{aligned} &\delta_s x^\mu = (\dot{x}^\mu/\eta\sqrt{\dot{x}^2})\, c, \\ &\delta_s c = 0, \qquad\qquad \delta_s^2 = 0 \\ &\delta_s \bar{c} = b, \\ &\delta_s b = 0, \end{aligned} \tag{53}$$

imply $\delta_s L_\eta = (d/dt)(mc)$ and the BRS Lagrangian is

$$\begin{aligned} \tilde{L} &= L_\eta + \delta_s[\bar{c}(\psi(\dot{x}^2) - b/2)] \\ &= \left(m + 2\frac{d\psi}{d\dot{x}^2}\bar{c}\dot{c}\right)\eta\sqrt{\dot{x}^2} + b\psi - \frac{b^2}{2}. \end{aligned} \tag{54}$$

Without loss of generality we can take $\psi = k\dot{x}^2 - h$, corresponding to the proper-time gauge fixings. The momenta and the Euler–Lagrange equations are

$$\begin{aligned} &p_\mu = (m + 2k\bar{c}\dot{c})(\dot{x}_\mu/\eta\sqrt{\dot{x}^2}) + 2kb\dot{x}_\mu, \\ &\pi_b = 0, \\ &\mathscr{P} = -2k\eta\sqrt{\dot{x}^2}\dot{\bar{c}} \quad \Rightarrow b\mathscr{P} = -(\eta\sqrt{p^2} - m)\bar{c} \\ &\bar{\mathscr{P}} = 0, \end{aligned} \tag{55}$$

$$\begin{aligned} &L_\mu = \frac{d}{d\tau}\left[(m + 2k\bar{c}\dot{c} + 2kb\eta\sqrt{\dot{x}^2})\frac{\dot{x}_\mu}{\eta\sqrt{\dot{x}^2}}\right] \doteq 0, \\ &L_b = k\dot{x}^2 - h - b \doteq 0, \\ &L_c = -2k\frac{d}{d\tau}(\eta\sqrt{\dot{x}^2}\dot{\bar{c}}) \doteq 0, \\ &L_{\bar{c}} = 2k\eta\sqrt{\dot{x}^2}\ddot{c} \doteq 0. \end{aligned} \tag{56}$$

Therefore, if we put $[m + 2k(k\dot{x}^2 - h)\eta\sqrt{\dot{x}^2}]\,(\dot{x}^\mu/\eta\sqrt{\dot{x}^2}) \doteq A^\mu$ with A^μ a timelike integration constant ($p^2 \doteq A^2$), we get $\eta\sqrt{\dot{x}^2} \doteq F(\eta\sqrt{A^2} - m, k, h)$, where F is the solution of the equation $\eta\sqrt{\dot{x}^2}(k\dot{x}^2 - h) \doteq (1/2k) \times (\eta\sqrt{A^2} - m)$. The solutions of Eqs. (56) are

$$\begin{aligned} &x^\mu(\tau) \doteq x^\mu(0) + F\frac{A^\mu}{\eta\sqrt{A^2}}\tau, \quad p^\mu \doteq A^\mu, \\ &b(\tau) \doteq k\dot{x}^2 - h \doteq kF^2 - h, \\ &c(\tau) \doteq c_0, \quad \mathscr{P}(\tau) \doteq -2kF\bar{c}_0, \\ &\bar{c}(\tau) \doteq \bar{c}_0. \end{aligned} \tag{57}$$

When $\eta\sqrt{p^2} \doteq \eta\sqrt{A^2} = m$ we get $F^2 \doteq h/k$, i.e., $\dot{x}^2 \doteq h/k$ and $b \doteq 0$, which are the extremals of L_η in the proper time gauges.

If one studies the Hessian matrix of $\tilde{L}_\eta$, one finds that one of its nonzero eigenvalues vanishes for the isolated value $b = 0$, reproducing the null eigenvalue of L_η. This phenomenon of variable rank of the Hessian matrix implies that for $b = 0$: (i) Eqs. (55) cannot be inverted to get $\dot{x}^2$ in terms of p^2 and $\bar{c}\dot{c}$; (ii) there is an induced primary constraint $\eta\sqrt{p^2} - m \approx 0$; (iii) the primary constraint $b\mathscr{P} + (\eta\sqrt{p^2} - m)\bar{c} \approx 0$ is absent. Moreover, we obtain $\bar{H}_c = \dot{x}\cdot p + \dot{b}\pi_b + \dot{c}\mathscr{P} + \dot{\bar{c}}\bar{\mathscr{P}} - \tilde{L}_\eta = b^2/2 - b(k\dot{x}^2 - h) + \eta\sqrt{\dot{x}^2}(\eta\sqrt{p^2} - m)$. Therefore, we need a generalized Legendre transformation[2] to define $\bar{H}_c$: $\eta\sqrt{\dot{x}^2} \mapsto \mu^1(\tau)$, where μ^1 is the induced Dirac multiplier of the induced primary constraint $\eta\sqrt{p^2} - m \approx 0$. The Dirac Hamiltonian is

$$\begin{aligned} \bar{H}_D = b^2/2 &- b\,[k(\mu^1(\tau))^2 - h\,] \\ &+ \mu^1(\tau)(\eta\sqrt{p^2} - m) + \mu^2(\tau)\pi_b + \mu^3(\tau) \\ &\times [b\mathscr{P} + (\eta\sqrt{p^2} - m)\bar{c}] + \mu^4(\tau)\bar{\mathscr{P}}, \end{aligned} \tag{58}$$

with $\mu^3(\tau) = 0$ for $b = 0$, because the corresponding constraint is absent. Then we get

$$\begin{aligned} &\dot{\pi}_b = \{\pi_b, \bar{H}_p\} = -b + k(\mu^1)^2 - h - \mu^3\mathscr{P} \approx 0, \\ &\dot{\bar{\mathscr{P}}} = \{\bar{\mathscr{P}}, \bar{H}_D\} = \mu^3(\eta\sqrt{p^2} - m) \approx 0, \\ &\frac{d}{d\tau}[b\mathscr{P} + (\eta\sqrt{p^2} - m)\bar{c}] \\ &\quad = \{b\mathscr{P} + (\eta\sqrt{p^2} - m)\bar{c}, \bar{H}_D\} \\ &\quad = \mu^2\mathscr{P} + (\eta\sqrt{p^2} - m)\mu^4 \approx 0, \quad \text{for} \quad b \neq 0. \end{aligned} \tag{59}$$

In the sector $b \neq 0$, $\eta\sqrt{p^2} - m \neq 0$ from Eqs. (59) we get $\mu^3 = 0$, $\mu^4 = -\mu^2\mathscr{P}/(\eta\sqrt{p^2} - m)$ and the secondary constraint $b - [k(\mu^1)^2 - h\,] \approx 0$, whose τ-constancy implies $\mu^2 = 2k\mu^1\dot{\mu}^1$. As the reality condition for the final Dirac Hamiltonian requires $\mu^4 = 0$, we get $\mu^2 = \dot{\mu}^1 = 0$, i.e., $\mu^1 = \text{const}$. The final result is: two pairs of 2nd-class constraints eliminating, b, π_b, $\bar{c}$, $\bar{\mathscr{P}}$, and, by using Dirac brackets, $\bar{H}^F_{D1} = \mu^1(\eta\sqrt{p^2} - m) - \frac{1}{2}[k(\mu^1)^2 - h\,]$. The Hamiltonian equations give $\dot{x}^\mu \doteq \mu^1 p^\mu/\eta\sqrt{p^2}$ and reproduce the Euler–Lagrange equations.

Instead, in the sector $b \approx 0$, $\eta\sqrt{p^2} - m \approx 0$, where $\mu^3 = 0$, the first two Eqs. (59) imply only $\mu^1 = \sqrt{h/k}$ and $\dot{b} \approx 0$ gives $\mu^2 = 0$. We now have a pair of 2nd-class constraints $b \approx 0$, $\pi_b \approx 0$, $\mu^4 = 0$ from the reality condition of the Dirac Hamiltonian, which turns out to be $\bar{H}^F_{D2} \equiv \sqrt{h/k}(\eta\sqrt{p^2} - m)$, and the two constraints $\bar{\mathscr{P}} \approx 0$, $\eta\sqrt{p^2} - m \approx 0$. These two constraints have been called 3rd-class in Ref. 2, because their Dirac multipliers are fixed ($\mu^1 = \sqrt{h/k}$, $\mu^4 = 0$), i.e., their conjugated variables satisfy deterministic 1st-order equations of motion: $\dot{x}^\mu \doteq \sqrt{h/k}\,p^\mu/\eta\sqrt{p^2}$ (proper-time gauge) and $\dot{\bar{c}} \doteq 0$: therefore, they are neither 1st- nor 2nd-class.

In both cases, from Eqs. (23) the conserved BRS charge is

$$\Omega = (2k\bar{c}\dot{c} + 2kb\eta\sqrt{\dot{x}^2})c = (\eta\sqrt{p^2} - m)c = \bar{\Omega}, \tag{60}$$

and we have $\bar{H}^F_{D1} = \{-\mu^1\mathscr{P}, \bar{\Omega}\} - \frac{1}{2}[k(\mu^1)^2 - h\,]$, $\bar{H}^F_{D2} = \{-\sqrt{h/k}\,\mathscr{P}, \bar{\Omega}\}$.

We conclude that the Lagrangians $L_\eta = \eta m\sqrt{\dot{x}^2}$ describe a single sheet of the mass hyperboloid and their BRS gauge fixings to the proper-time gauges imply a nontrivial Kato–Ogawa approach. (In Sec. II, we had two sectors but no problems of variable rank and of projectability to phase space.) Let us remark that L_η has the associated Dirac Hamiltonian $\bar{H}_{\rm D} = \rho(\tau)(\eta\sqrt{p^2} - m)$ and it coincides with $\bar{H}^{F}_{\rm D2}$ for $\rho(\tau) = \sqrt{h/k}$: since there is only one constraint, this value of $\rho(\tau)$ (proper-time gauge) can be obtained for instance with the Dirac gauge fixing $p\cdot x/\eta\sqrt{p^2} - \sqrt{h/k}\,\tau \approx 0$.

This example shows that covariant gauge fixings may also be implemented with a Kato–Ogawa approach, but at the price of a more complex Hamiltonian theory. The advantage of the BFV approach is to have a simple Hamiltonian description, but with a doubling of the number of constraints and ghosts.

IV. OTHER ASPECTS OF THE BRS THEORY

Let us now try to find a connection among the previous BRS approaches, the multitemporal description of models with 1st-class constraints,[1] the Hamiltonian Konstant–Sternberg[21] algebraic point of view on the BRS approach, its differential geometry reinterpretation given by Loll,[22] and finally the Bonora–Cotta Ramusino[23,6] interpretation of the ghosts at the Lagrangian level. This would permit us to understand both the static kinematical aspects of BRS theory (formal interpretation of the ghosts, of the BRS invariance, and of the extended phase space) and the dynamical ones given by either the Euler–Lagrange or the Hamiltonian equations for the ghosts, whose solutions identify the residual gauge freedom left by the chosen allowed Lagrangian gauge fixing.

Let us suppose to have a system described by a singular Lagrangian $L(q^i,\dot{q}^i)$, $i = 1,...,n$, whose Hessian matrix has m null eigenvalues. Let L be quasiinvariant under m Nœther transformations $\delta_A q^i = \epsilon^A(\tau)\,{}_A\xi^i_0(q,\dot{q})$, where ${}_A\xi^i_0$ are the null eigenvectors of the Hessian matrix. This implies the existence of m primary 1st-class constraints $\bar{\phi}_A \approx 0$ in phase space. For the sake of simplicity we have chosen a system without secondary constraints [$\delta_A q^i$ depends only on $\epsilon^A(\tau)$ and not on its τ-derivatives[1,2]]. If the null eigenvectors ${}_A\xi^i_0$ have been chosen orthonormal (${}_A\xi^i_0 = {}_A\xi^{i(0)}_0$) we have ${}_A\xi^{i(0)}_0 = \partial\bar{\phi}^{(0)}_A/\partial p_i$, for a certain functional form of the constraints. It turns out that the ${}_A\xi^i_0$'s can be chosen orthonormal only locally in general and that with the local form $\bar{\phi}^{(0)}_A$ one locally gets[2] $\{\bar{\phi}^{(0)}_A,\bar{\phi}^{(0)}_B\} = 0$ (local abelianization[10]). Let us assume that there exists a global form for the nonorthonormalized ${}_A\xi^i_0$ such that the associated constraints $\bar{\phi}_A$, obtained via the second Nœther theorem,[1,2] are globally defined, so that they globally define the constraint submanifold $\bar{\gamma}\subset T^*Q$ of phase space. Let us also assume that these globally defined constraints $\bar{\phi}_A$ satisfy $\{\bar{\phi}_A,\bar{\phi}_B\} = -C^C_{AB}\bar{\phi}_C$, where C^C_{AB} are the structure constants of a Lie algebra g. The associated Hamiltonian vector fields $\bar{X}_A = \{.,\bar{\phi}_A\}$ then satisfy $[\bar{X}_A,\bar{X}_B] = C^C_{AB}\bar{X}_C$. The set of all the possible functional forms of the constraints, all of them locally defining the same submanifold $\bar{\gamma}$, are connected via the theory of function groups.[24] The transition from one functional form to another one cannot be realized in general with a canonical transformation in phase space, but it can be done with a supercanonical transformation in the extended phase space.[10]

The change in the functional form of the constraints is a reflex of the parametrization of the gauge transformations $\delta_A q^i = \epsilon^A(\tau)f_A(q,\dot{q})\,{}_A\xi^{i(0)}_0(q,\dot{q}) = \epsilon^A(\tau)\,{}_A\xi^i_0(q,\dot{q})$, where the functions f_A generalize the f of Sec. II. For each choice of the f_A let us put $\epsilon^A(\tau) = \rho^A c^A(\tau)$ and let us study the BRS transformations. Let $\{f_A\}$ denote the set of functions f_A for which $\delta^2_s = 0$: to it will correspond a set of functional forms $\{\bar{\phi}_A\}$ of the constraints that are compatible with the classical BRS theory. This means that for all these functional forms, the gauge algebra hypothesis is globally satisfied. Let the previously discussed global constraints belong to this set and let us consider only them in what follows.

Let us consider those BRS Lagrangians, with allowed Lagrangian gauge fixings in the sense of Sec. II, which do not contain terms $d(\bar{c}_A\eta^A)/d\tau$, so to get a phase space Kato–Ogawa theory (the same considerations could be made for the BFV case with a doubling of the constraints). Therefore, one gets fermionic constraints implying that the momenta $\mathscr{P}_A$ are proportional to the antighosts $\bar{c}_A$: from now on we will speak only of $\mathscr{P}_A$ and not of $\bar{c}_A$ [in the BFV case the role of the ghosts c^A is taken by $(c^A,\bar{\mathscr{P}}^A)$ and the role of the $\mathscr{P}_A$ by $(\mathscr{P}_A,\bar{c}_A)$]. Let us assume that the BRS charge has the standard form $\bar{\Omega} = c^A\bar{\phi}_A - \frac{1}{2}C^C_{AB}c^Ac^B\mathscr{P}_C$, so that $\delta_s q^i = c^A\{q^i,\bar{\phi}_A\}$ satisfies $\delta^2_s q^i = 0$.

Our hypotheses until now amount to saying that there exist[21] a linear action of the Lie algebra g, δ: $g\to F(T^*Q) = C^\infty(T^*Q)$, into the set of C^∞ functions $\bar{f}(q,p)$ on the phase space T^*Q. If e_A is a basis for $g([e_A,e_B] = C^C_{AB}e_C)$, we have δ: $e_A\mapsto\bar{\phi}_A$, where $\bar{\phi}_A$ are the globally defined constraints; the dual basis for the dual g^* of the Lie algebra g will be denoted with e^A and $e^A(e_B) = \delta^A_B$. Moreover, we have a Hamiltonian right action φ: $g\to\chi(T^*Q)$, where $\chi(T^*Q)$ are the vector fields over T^*Q, such that φ: $e_A\mapsto\bar{X}_A = \{.,\bar{\phi}_A\}$. The $\bar{X}_A$ are then called the fundamental vector fields. Let us assume[25,26] that the action φ is: (i) effective (i.e., it is an isomorphism onto its image); (ii) free (i.e., the $\bar{X}_A$'s are independent and without zeroes); (iii) foliating (i.e., T^*Q is foliated with m-dimensional gauge orbits, all diffcomorphic; only the gauge orbits $\bar{\Gamma}_p$ of $\bar{\gamma}\subset T^*Q$ are physically relevant); (iv) regular (i.e., the gauge orbits are regular submanifolds of T^*Q). Then the vector field $\bar{X}_A$ restricted to a gauge orbit $\bar{\Gamma}_p$ become a basis for its tangent vector fields.

Let us remark that with a singular Lagrangian only the primary constraint submanifold $\bar{\gamma}$ of T^*Q is identified by the Legendre transformation. The Dirac–Bergmann description of constrained systems is based on making an extrapolation of all the relevant quantities from $\bar{\gamma}$ to a small neighborhood of $\bar{\gamma}$ in T^*Q. Instead, in the previous hypotheses one is assuming the possibility of doing such an extrapolation to the whole T^*Q. The Dirac–Bergmann approach together with its BRS extensions would need to be reformulated in an intrinsic way by using only the manifold $\bar{\gamma}$, which is presymplectic. Since this will be done elsewhere, let us go on with this framework using the whole T^*Q like in Ref. 21.

The next problem is to integrate the local action φ of g to a Hamiltonian right group action. A necessary condition is the completeness of the vector fields $\bar{X}_A$. In general, when the integration is possible, one identifies only a local Lie group G. For the sake of simplicity let us assume that our system is such that the φ action can be globally integrated to a Hamiltonian right action of a well-defined Lie group G with Lie algebra g on T^*Q: i.e., there exist $\psi: G\to \mathrm{Diff}(T^*Q)$, associating to each group element G_A a diffeomorphism of T^*Q, implemented via Poisson brackets. This implies that all the gauge orbits are homogeneous spaces, on which G acts transitively, and are therefore diffeomorphic to the group manifold of G. Moreover, let us assume that G acts freely, that the reduced phase space (space of the gauge orbits) $\bar{\gamma}_R = \bar{\gamma}/G$ is a manifold and that π: $\bar{\gamma}\to\bar{\gamma}_R$ is a principal G bundle. Also let G be connected, simply connected and compact so that the exponential map from g to G spans the whole G: the associated exponential map from $\varphi(g)$ to $\psi(G)$ will allow us to reconstruct the whole gauge orbits starting from a point of each one of them. Therefore, $\bar{\gamma}$ (and also T^*Q) is foliated with these gauge orbits and in this framework using T^*Q one can also introduce the momentum map[21] μ: $T^*Q\to g^*, \mu: z\in T^*Q\to e^A\bar{\phi}_A(z)$, so that $\bar{\gamma}=\mu^{-1}(0)$: i.e., the constraint submanifold is the inverse image under μ of the coadjoint orbit corresponding to the origin of g^*.

Let us first study the multitemporal approach[1,27] to such a system. Besides the primary constraints $\bar{\phi}_A$, there will be a canonical Hamiltonian $\bar{H} = \dot{q}^i p_i - L$, which in general will satisfy $\{\bar{H},\bar{\phi}_A\} = \bar{V}^B_A(q,p)\bar{\phi}_B \approx 0$. For the sake of simplicity, let us assume that $\{\bar{H},\bar{\phi}_A\} = 0$, when the $\bar{\phi}_A$'s are the globally defined constraints. The case $\{\bar{H},\bar{\phi}_A\} = \bar{V}^B_A\bar{\phi}_B$ could be treated from the algebraic point of view in the same way by considering a Lie algebra g' with structure constants C^C_{AB} and $C^B_{0A} = V^B_A$ and an extra generator $e_0\to\bar{X}_0 = \{.,\bar{H}\}$ besides the $e_A\mapsto\bar{X}_A$. The Dirac Hamiltonian is $\bar{H}_D = \bar{H} + \Sigma_A\lambda^A(\tau)\bar{\phi}_A$, where the $\lambda^A(\tau)$'s are the Dirac multipliers and the Hamilton equations are

$$\frac{d}{d\tau}\bar{f}(q,p) \doteq \{\bar{f},\bar{H}_D\} = \bar{X}_0\bar{f} + \sum_A \lambda^A(\tau)\bar{X}_A\bar{f}. \tag{61}$$

For fixed $\lambda^A(\tau)$, Eqs. (61), with the initial data given by a point $z^\alpha = (q^i,p_i)\in\bar{\gamma}$ lying on the gauge orbit $\bar{\Gamma}_0\subset\bar{\gamma}$, describe its evolution as composed of two parts: an evolution on the gauge orbit $\bar{\Gamma}_0$, generated by $\Sigma_A\lambda^A(\tau)\bar{X}_A$, plus a deterministic evolution from the gauge orbit $\bar{L}_0$ to another gauge orbit, generated by $\bar{X}_0$. Since each gauge orbit $\bar{\Gamma}$ is diffeomorphic to the Lie group G, whose group manifold can be described with coordinates τ^A, the generator $\Sigma_A\lambda^A(\tau)\bar{X}_A$ is the counterpart on $\bar{\Gamma}$ of the generator of an one-parameter subgroup of G, identified by a set of functions $\tau^A(\tau)$ [$\tau^A_0 = \tau^A(0)$ denote the point in G corresponding to z^α in $\bar{\Gamma}_0$].

In G we can introduce the Maurer–Cartan left-invariant one-forms $\theta^A = A^A_B(\tau^C)d\tau^B$ and their dual left-invariant vector fields $Y_A = B^B_A(\tau^C)(\partial/\partial\tau^B)$, where $i_{Y_A}\theta^B = \delta^B_A$ implies $A^A_B B^B_C = \delta^A_C$. If E is the identity in G, with coordinates $\tau^A = 0$, we have $Y_A|_E = e_A$, where e_A are the generators of g, and $\theta^A|_E = e^A$, with e^A the generators of g^* (g and g^* are identified with T_EG and T^*_EG, respectively). We have

$$[Y_A,Y_B] = C^C_{AB}Y_C \Rightarrow A^E_A\frac{\partial A^D_B}{\partial\tau^E} - A^E_B\frac{\partial A^D_A}{\partial\tau^E} = C^C_{AB}A^D_C, \tag{62}$$

$$d\theta^A = -\tfrac{1}{2}C^A_{BC}\theta^B\wedge\theta^c. \tag{63}$$

Both Eqs. (62) (63) are called Maurer–Cartan equations.

Therefore, to assign the Dirac multipliers $\lambda^A(\tau)$ is equivalent to assigning some set of functions $\tau^A(\tau)$ and the connection is given by

$$\lambda^A(\tau) = A^A_B(t^c(\tau))\frac{d\tau^B(\tau)}{d\tau} \Rightarrow \lambda^A(\tau)d\tau = \theta^A|_{\tau^c=\tau^c(\tau)}. \tag{64}$$

This suggests that the Dirac–Bergmann theory with the associated Hamilton equations (61), should be rephrased in the following way: the coordinates $q^i(\tau)$, $p_i(\tau)$, which for each value of τ contain the gauge degrees of freedom associated to the gauge transformations related to the gauge group G, must be replaced by functions $q^i(\tau,\tau^C),p_i(\tau,\tau^C)$ which depend simultaneously on an evolution parameter τ for $\bar{X}_0 = \{.,\bar{H}\}$ and on the group manifold parameters τ^C. The Hamilton equations (61), with arbitrary $\lambda^A(\tau)$'s, are then replaced by the multitemporal (in τ and τ^C) Hamilton equations

$$\frac{\partial}{\partial\tau}\bar{f}(q,p) \doteq \{\bar{f},\bar{H}\} = \bar{X}_0\bar{f},$$
$$Y_A\bar{f}(q,p) = A^B_A(\tau^c)\frac{\partial}{\partial\tau^B}\bar{f} \doteq \{\bar{f},\bar{\phi}_A\} = \bar{X}_A\bar{f}, \quad A = 1,\dots,m. \tag{65}$$

The second set of Eqs. (65) are just the Lie equations for the Lie group G, acting as a transformation group on $\bar{\gamma}\in T^*Q$ via a symplectic right action.[28,26] This means that the second set of Eqs. (65), when restricted to a gauge orbit $\bar{\Gamma}$ with given initial conditions on it, allow the reconstruction of $\bar{\Gamma}$ by means of their integration with those initial conditions. The integrability conditions of Eqs. (65) are just the Maurer–Cartan equations (62) plus their analogs on $\bar{\gamma}$, i.e., $[\bar{X}_A,\bar{X}_B] = C^C_{AB}\bar{X}_C$:

$$[Y_A,Y_B]\bar{f} = C^C_{AB}Y_C\bar{f} \doteq [\bar{X}_A,\bar{X}_B]\bar{f} = C^c_{AB}\bar{X}_c\bar{f}, \tag{66}$$
$$0 = \left[Y_A,\frac{\partial}{\partial\tau}\right]\bar{f} \doteq [\bar{X}_A,\bar{X}_0]\bar{f} = 0.$$

The second set of Eqs. (65) carry the information which is missing in the Euler–Lagrange equations for L: the undetermination in the extremals of L due to the gauge transformations is not totally arbitrary, but is connected to the Lie group G associated to the gauge freedom.

When a Lagrangian gauge fixing is added to L, the resulting "residual gauge freedom" becomes a restriction in the functional form of the Dirac multipliers $\lambda^A(\tau)\mapsto\tilde{\lambda}^A(\tau)$, as presented in Sec. I. This restriction may change from point to point of $\bar{\gamma}\in T^*Q$ if $\tilde{\lambda}^A = \lambda^A(\tau|q,p)$ [remember $\tilde{\epsilon} = \tilde{\epsilon}(\tau|q,\dot{q})$]. This means that instead of spanning all the possible one-parameter subgroups of G one is restricted to span only a special subset of them identified by the residual gauge freedom. This subset could change either from a gauge orbit to the other in a smooth way, if $\tilde{\lambda}^A = \tilde{\lambda}^A(\tau|\bar{\Gamma})$, or also from a point to another one in the same gauge orbit, if $\tilde{\lambda}^A = \tilde{\lambda}^A(\tau|q,p)$.

When we add an allowed set of Dirac gauge fixings $\bar{\chi}_A \approx 0$, the equations $\dot{\bar{\chi}}_A \doteq \{\bar{\chi}_A, \bar{H}_D\} \approx 0$ determine the $\lambda^A(\tau)$'s and therefore a unique one-parameter subgroup is chosen: by means of Eqs. (64) we see that Eqs. (61) with these $\lambda^A(\tau)$'s are the corresponding "one-time" theory and it is recovered from the linear combination of Eqs. (65) corresponding to the following decomposition:

$$\frac{d}{d\tau} = \frac{\partial}{\partial\tau} + \lambda^A(t) Y_A \Big|_{\tau^c = \tau^c(\tau)} = \frac{\partial}{\partial\tau} + \frac{d\tau^A(\tau)}{d\tau} \frac{\partial}{\partial\tau^A}\Big|_{\tau^c = \tau^c(\tau)}. \qquad (67)$$

If for instance the Lagrangian gauge fixing is such to give a family of residual $\tilde{\lambda}^A(\tau)$, which can be interpreted as the assignment of functions $\tau^A = \tau^A(\rho^a)$, $A = 1,...,m$, $a = 1,...,k < m$, where ρ^a are the coordinates of a subgroup G' of G, the $(m+1)$-times Eqs. (65) are restricted everywhere on $\bar{\gamma}$ to the following $(k+1)$-times equations:

$$\frac{\partial}{\partial\tau}\bar{f} \doteq \bar{X}_0 \bar{f},$$

$$Z_a \bar{f} = D^b_a(\rho) \frac{\partial}{\partial\rho^b}\bar{f} = D^b_a(\rho)\frac{d\tau^A(\rho)}{d\rho^b} A^B_A(\tau(\rho)) \frac{\partial}{\partial\tau^B}\Big|_{\tau=\tau(\rho)} \bar{f}$$

$$\doteq D^b_a(\rho)\frac{d\tau^A(\rho)}{d\rho^b}\bar{X}_A \bar{f} \underset{\text{def}}{=} \bar{X}_a \bar{f}, \qquad (68)$$

$$\bar{X}_a = \bar{g}^A_a \bar{X}_A, \quad \bar{g}^A_a \doteq D^b_a \frac{d\tau^A(\rho)}{d\rho^b},$$

$$[Z_a, Z_b] = C'^c_{ab} Z_c, \quad [\bar{X}_a, \bar{X}_b] = C'^c_{ab}\bar{X}_c,$$

$$\bar{X}_a = \{\cdot, \bar{g}^A_a(q,p)\bar{\phi}_A\} \approx \bar{g}^A_a \bar{X}_A.$$

Here, the $\bar{g}^A_a(q,p)$ are those suitable functions on phase space which select the generators $\bar{X}_a$ of the Lie algebra g' of G', which is a generalized Lie subalgebra of g (it is a Lie subalgebra of g when $g^A_a = \text{const}$). In this case, the functions $\bar{f}'$ having vanishing Poisson brackets with the linear combinations of the $\bar{\phi}_A$'s functionally independent from the $\bar{\phi}_a$'s, are not yet observables. The observables are $\bar{f}_0$, $\{\bar{f}_0, \bar{\phi}_A\} \approx 0$, $A = 1,...,m$, because there is still the residual gauge freedom connected to the generators $\bar{X}_a$. The $\bar{f}_0$ are recovered from the $\bar{f}'$ by going to the quotient with respect to these residual gauge transformations.

However, in general, it is difficult to recast the effect of a Lagrangian gauge fixing in this multitemporal approach, because the corresponding equations of the type of Eqs. (68) could vary in a smooth way from point to point of $\bar{\gamma} \in T^*Q$.

It is here that the BRS approach becomes really important. First of all it selects the functional forms of the constraints compatible with $\delta_s^2 = 0$ (see Sec. II). Then it replaces Eqs. (68) with the BRS Hamilton equations on the extension of T^*Q, which, when restricted to $\bar{\gamma}$, contain the information on the residual gauge freedom hidden in the ghosts c^A: $\rho c^A(\tau) \doteq \tilde{\epsilon}(\tau)$. The conserved BRS charge $\bar{\Omega}$ is then used to find the BRS observables, i.e., those functions $\bar{F}$ on the extended phase space satisfying $\{\bar{F}, \bar{\Omega}\} = 0$. Each $\bar{F}$ is defined modulo the trivial BRS observables $\{\bar{R}, \bar{\Omega}\}$: this is the reflex of the residual gauge freedom in this approach. The real observables $\bar{f}_0(\{\bar{f}_0, \bar{\phi}_A\} = 0$, $A = 1,...,m)$ are recovered as the equivalence classes $(\bar{F} + \{\bar{R}, \bar{\Omega}\})$ for $\bar{F}$ with zero ghost number $N_g = c^A \mathscr{P}_A$ $(\{c^A, N_g\} = c^A, \{\mathscr{P}_A, N_g\} = -\mathscr{P}_A, \{\bar{\Omega}, N_g\} = \bar{\Omega})$, i.e., by making the quotient of the BRS observables with respect to the residual gauge transformations. This interpretation is consistent with the fact that $\bar{\Omega}$ is the generator of the global extended Nœther transformations for the BRS Lagrangian and, as noted in Secs. II and III, this is the global Nœther symmetry englobing the residual gauge transformations via $\rho c^A(\tau) \doteq \tilde{\epsilon}(\tau)$. Further support to this interpretation will come from what follows.

This intuitive picture is at the basis of the Konstant–Sternberg approach[21] to the Hamiltonian BRS theory. Their algebraic point of view is centered on the fact that we are interested more on the observables $\bar{f}_0$ rather than on the Hamilton equations (61) and (65). These observables are the functions defined on the reduced phase space $\bar{\gamma}_R = \bar{\gamma}/G$, which is a symplectic manifold under our hypotheses, with Poisson brackets $\{.,.\}_R$. Since the canonical Hamiltonian $\bar{H}$ is an observable, the only relevant Hamilton equations are

$$\frac{d}{d\tau}\bar{f}_0 \doteq \{\bar{f}_0, \bar{H}\}_R, \quad \{\bar{f}_0, \bar{\phi}_A\} = 0. \qquad (69)$$

Therefore, Konstant–Sternberg abandon the detailed study of the dynamics on $\bar{\gamma} \subset T^*Q$ and only concentrate on the functions $\bar{f}(q,p)$ on T^*Q. They develop a method based on homology and cohomology to extract the observables $\bar{f}_0$ from the $\bar{f}$'s (see Refs. 29–31 for other approaches to classical BRS cohomology). They interpret the ghosts c^A, as a basis, e^A, for the dual $g*$ of the Lie algebra g of G and the momenta $\mathscr{P}'_A$ as a basis, e_A, for g. Then they define the Grassmann (exterior) algebras $\Lambda g*, \Lambda g$ associated to $g*$ and g, respectively: $g*_1 \Lambda \cdots \Lambda g*_p \in \Lambda^p g*, g_1 \Lambda \cdots \Lambda g_p \in \Lambda^p g$. They consider the complex $\Lambda g* \otimes \Lambda g \otimes \mathscr{F}(T^*Q)$. By using the linear action δ: $e_A \mapsto \bar{\phi}_A$ of g they define a "boundary operator" δ: $\Lambda^p g \otimes \mathscr{F}(T^*Q) \mapsto \Lambda^{p-1} g \otimes \mathscr{F}(T^*Q)$, $\delta^2 = 0$, and show that the associated zeroth homology group is equivalent to the quotient of $\mathscr{F}(T^*Q)$ with respect to its subspace formed by all the functions $\bar{f}$ vanishing on the constraint manifold $\bar{\gamma} \subset T^*Q$: in this way we get a characterization of $\mathscr{F}(\bar{\gamma})$. Then they introduce a "coboundary operator" d: $\Lambda^p g* \otimes \mathscr{F}(T^*Q) \mapsto \Lambda^{p+1} g* \otimes \mathscr{F}(T^*Q)$, $d^2 = 0$, whose zeroth cohomology group identifies the subspace of $\mathscr{F}(T^*Q)$ containing all the G-invariant functions on T^*Q. The BRS operator is the "coboundary operator" $D = d + (-)^p 2\delta$ (when acting on $\Lambda^p g$), $D^2 = 0$, so that, due to the combined effects of d and δ, its zeroth cohomology group identifies the G-invariant functions present in $\mathscr{F}(\bar{\gamma})$, i.e., the observables $\bar{f}_0$. Then after a canonical identification of $\Lambda g* \otimes \Lambda g$ with $\Lambda(g* \oplus g)$, they succeed to introduce a super-Poisson bracket such that $\{c'^A, \mathscr{P}'_B\} = 2\delta^A_B$, $\{c'^A, c'^B\} = \{\mathscr{P}'_A, \mathscr{P}'_B\} = 0$: at this stage, the elements of $g* \oplus g$ have become Grassmann variables, i.e., something like one-forms. Finally, they give the representation $D=\{.,\overline{\Omega}'\}$, where $\overline{\Omega}'$ is the BRS charge.

With our notations, this amounts to the following redefinition:

$$\begin{aligned} &c^A\to c'^A=-c^A,\\ &\mathcal{P}_A\to\mathcal{P}'_A=2\mathcal{P}_A,\\ &\{c^A,\mathcal{P}_B\}=-\delta^A_B\to\{c'^A,\mathcal{P}'_A\}=2\delta^A_B,\\ &\overline{\Omega}=c^A\bar{\phi}_A-\tfrac{1}{2}C^c_{AB}c^Ac^B\mathcal{P}_c\\ &\quad\to -c'^A(\bar{\phi}_A+\tfrac{1}{4}C^c_{AB}c'^B\mathcal{P}'_c), \end{aligned}\tag{70}$$

so that we get ($z^\alpha=(q^i,p_i)$):

$$\begin{aligned} \delta_s z^\alpha&=\{z^\alpha,\overline{\Omega}\}=\{z^\alpha,\bar{\phi}_A\}c^A=(\overline{X}_Az^\alpha)c^A\\ &=\delta_A z^\alpha c^A=-\delta_A z^\alpha c'^A,\\ \delta_s c^A&=-\{c^A,\overline{\Omega}\}=-\tfrac{1}{2}c^A_{BC}c^Bc^c\to\delta_s c'^A\\ &=-\{c'^A,\overline{\Omega}\}=\tfrac{1}{2}C^A_{BC}c'^Bc'^c,\\ \delta_s\mathcal{P}_A&=-\{\mathcal{P}_A,\overline{\Omega}\}\\ &=\bar{\phi}_A-C^c_{AB}c^B\mathcal{P}_c\to\delta_s\mathcal{P}'_A\\ &=-\{\mathcal{P}'_A,\overline{\Omega}\}=2\bar{\phi}_A+C^c_{AB}c'^B\mathcal{P}'_c. \end{aligned}\tag{71}$$

Due to the second line of Eqs. (71) one says that the ghosts satisfy the Maurer–Cartan equations by analogy with Eqs. (63).

In a subsequent paper by Loll[22] an attempt is made to reinterpret the Konstant–Sternberg algebraic approach from a differential geometry point of view, which is more desirable for the study of the global properties of the system and for the connection with the Hamilton equations. Loll's construction suggests the following interpretation of the extended phase space $S_{\rm ext}$ with coordinates $(q^i,p_i,c'^A,\mathcal{P}'_A)$: The supermanifold $S_{\rm ext}$ is a trivial vector bundle having T^*Q as base space, $\Lambda(g^*\oplus g)$ as fiber and G as structure group, associated to the trivial vector bundle $T^*Q\times(g^*\oplus g)$. Here, $S_{\rm ext}$ is a symplectic manifold admitting the Konstant–Sternberg super-Poisson bracket. The group G acts on the fiber $\Lambda(g^*\oplus g)$ by means of its adjoint action on g and its coadjoint action on g^*[25,26]:

$$\mathrm{AD}_V(a^*,b)=(\mathrm{Ad}^*_{V^{-1}}\,a^*,\mathrm{Ad}_V\,b),\quad V\in G,a^*\in g^*,b\in g. \tag{72}$$

The associated frame bundle to $S_{\rm ext}$ (i.e., the bundle of all frames over $S_{\rm ext}$, defining its local coordinates) has Osp $(m,m,2n)$ as its structure group. This would explain the relevance of the Osp's groups even when the anti-BRS charge is not a constant of the motion. The action of G on $g^*\oplus g$ is symplectically realized by a means of a subgroup of the $O(m,m)$ transformations contained in Osp$(m,m,2n)$. In this way one is also choosing the subgroup Sp$(2n)$ of Osp$(m,m,2n)$ as the relevant subgroup of the canonical transformations of T^*Q. In Osp $(m,m,2n)$ there are also generators performing infinitesimal changes of the functional form of the constraints. The triviality condition is assumed because only in this case one can do a separate global quantization of the even and odd variables, obtaining the tensor product of the associated Hilbert spaces.

The trivial vector bundle, $S_{\rm ext}$ over T^*Q, is derived by another nontrivial vector bundle $S'_{\rm ext}$ over T^*Q/G, which is well defined when the action of G is foliating the whole T^*Q and not only $\bar{\gamma}$ (so that T^*Q/G contains $\bar{\gamma}_R=\bar{\gamma}/G$), always with fiber $\Lambda(g^*\oplus g)$ and structure group G. The cross sections of $S'_{\rm ext}$ are functions on T^*Q with values in $\Lambda(g^*\oplus g)$ and therefore belong to the Konstant–Sternberg complex $\mathcal{F}(T^*Q)\otimes\Lambda(g^*\oplus g)$. The key observation of Loll is that a function $\overline{F}$ of the complex is compatible with the projection $T^*Q\to T^*Q/G$ if it is equivariant[25,26] with respect to the action (72) of G on $S'_{\rm ext}$: if s denotes the whole set of variables $\overline{F}$ depends upon and $V\in G$, this means $[U(V)\psi](s)=\psi(sV)=\mathrm{AD}_{V^{-1}}\psi(s)$ and for V near the identity E of G one gets $\{\psi(s),\bar{\phi}_A\}=-\{\psi(s),\overline{H}_A\}$. Here, $\overline{X}_{H_A}=\{.,\overline{H}_A\}$ is the infinitesimal generator for the symplectic realization of the action AD and $[\overline{X}_{H_A},\overline{X}_{H_B}]=C^C_{AB}\overline{X}_{H_C}$. Then Loll looks for the subclass of the BRS-observables which satisfy the equivariance condition: she claims that they are independent from the $\mathcal{P}_A$'s, but it is not clear to us which BRS charge is used by her.

Therefore, let us go back to our notations and let us look for the generators $\overline{X}_{H_A}$ realizing the AD action. It turns out that the following $\overline{H}_A$ have the right properties:

$$\begin{aligned} &\overline{H}_A=-C^C_{AB}c^B\mathcal{P}_c,\\ &\{\bar{\phi}_A+\overline{H}_A,\bar{\phi}_B+\overline{H}_B\}=-C^C_{AB}(\bar{\phi}_c+\overline{H}_c), \end{aligned}\tag{73}$$

so that the equivariance condition is $\{\overline{F},\bar{\phi}_A+\overline{H}_A\}=0$ and our BRS charge becomes

$$\begin{aligned} \overline{\Omega}&=c^A(\bar{\phi}_A+\overline{H}_A)+\tfrac{1}{2}C^{C}_{AB}c^Ac^B\mathcal{P}_C\\ &=c^A(\bar{\phi}_A+\overline{H}_A)-\tfrac{1}{2}c^A\overline{H}_A. \end{aligned}\tag{74}$$

It seems reasonable to identify Loll's charge with $c^A(\bar{\phi}_A+\overline{H}_A)$, because for equivarient functions one gets that the equation

$$\{\overline{F},c^A(\bar{\phi}_A+\overline{H}_A)\}=\{\overline{F},c^A\}(\bar{\phi}_A+\overline{H}_A)=0, \tag{75}$$

has for solution $\mathcal{P}_A$-independent equivariant functions.

Instead, let us investigate which is the restriction imposed by the equivariance condition on the BRS observables by using the BRS charge of Eqs. (74):

$$\begin{aligned} \{\overline{F},\overline{\Omega}\}&=\{\overline{F},c^A\}(\bar{\phi}_A+\overline{H}_A)-\tfrac{1}{2}\{\overline{F},c^A\overline{H}_A\}\\ &=\{\overline{F},c^A\}\bar{\phi}_A+\tfrac{1}{2}C^C_{AB}c^Ac^B\{\overline{F},\mathcal{P}_c\}=0. \end{aligned}\tag{76}$$

Since these equations have to hold both on $T^*Q(\bar{\phi}_A\neq 0)$ and on $\bar{\gamma}(\bar{\phi}_A=0)$, Eqs. (76) imply (strictly speaking $\{\overline{F},\mathcal{P}_A\}$ is proportional to the constraints, but we can choose a prolongation of $\overline{F}_{|\bar{\gamma}}$ to T^*Q which is c^A independent)

$$\{\overline{F},c^A\}=\{\overline{F},\mathcal{P}_A\}=0\Rightarrow\{\overline{F},\overline{H}_A\}=0,\quad \overline{F}=\bar{f}(q,p), \tag{77}$$

so that the equivariance condition implies

$$\{\overline{F}=\bar{f}(q,p),\bar{\phi}_A\}=0\Rightarrow\overline{F}=\bar{f}=\bar{f}_0. \tag{78}$$

It turns out that the equivariance condition selects among the BRS observables the real ones $\bar{f}_0$, which are also the final result of the Konstant–Sternberg approach. Therefore, the equivariance condition realizes the passage to the quotient with respect to the trivial BRS observables (the residual gauge freedom) without the necessity to require by hand the vanishing of the ghost number of $\bar{F}$.

The equivariance condition is a natural requirement due to the identification $c^A \mapsto e^A \in g^*, \mathcal{P}_A \mapsto e_A \in g$ proposed by Konstant–Sternberg, so that the action of G on them must be AD. Since $\{\bar{\phi}_A + \bar{H}_A, \bar{\Omega}\} \neq 0$, we see that the trivial BRS observables are not equivariant and that $\bar{\phi}_A + \bar{H}_A$ is not a BRS observable.

Let us now go back to the multitemporal equations (65), whose second set implies that the original Nœther transformations $\delta_A q^i$ are a particular case of the phase space general gauge transformations

$$\delta_G \bar{f}(q,p) = \sum_A \epsilon^A(\tau) \bar{X}_A \bar{f} \doteq \sum_A \epsilon^A(\tau) Y_A \bar{f}. \tag{79}$$

Now under the substitution $\epsilon^A \mapsto \rho c^A(\tau)$, we get $c^A(\tau)\bar{X}_A \doteq c^A(\tau) Y_A$ which becomes $e^A(\tau)\bar{X}_A \doteq e^A(\tau) Y_A$ with the Konstant–Sternberg identification $c^A \mapsto e^A \in g^*$. In accord with Loll's construction on each point $z = (q,p)$ of $\bar{\gamma} \subset T^*Q$ is attached a copy of g^* and the multitemporal equations say that the generator of the gauge transformations, $e^A(\tau)\bar{X}_A$ (which, in general, changes with the time τ associated to the deterministic evolution generated by $\bar{H}$), is equal to $e^A(\tau) Y_A$, the corresponding generator on the group manifold of G (which is diffeomorphic to the gauge orbit containing z, with z considered as the origin of the coordinate system τ^A). Since $\bar{\gamma} \to \bar{\gamma}_R = \bar{\gamma}/G$ is a principal G-bundle under our hypotheses, Eq. (79) describe only a subgroup of the infinite-dimensional group $\mathcal{G}$ of the gauge transformations[23,32] [the concept of connection, i.e., the definition of the horizontal vectors in each point $z \in \bar{\gamma}$, is defined by looking for a $2(n-m)$-dimensional symplectic basis of observables around z and connecting these bases in a smooth way; the vertical vectors are $\bar{X}_A|_z$]. Here $\mathcal{G} = \Gamma(\mathrm{Ad}\,\bar{\gamma})$, i.e., the gauge transformations are identified with the cross sections of the bundle $\mathrm{Ad}\bar{\gamma} = \bar{\gamma} \times_G G$, associated to $\bar{\gamma} \mapsto \bar{\gamma}_R$, with fiber G and with G acting on itself with the adjoint action. For an element of $\mathcal{G}$, Eq. (79) is replaced by

$$\delta_G \bar{f}(z) = \sum_A \epsilon^A(\tau|\bar{\Gamma})(\bar{X}_A \bar{f})(z) \doteq \sum_A \epsilon^A(\tau|\bar{\Gamma}) Y_A \bar{f}(z(\tau)), \tag{80}$$

with ϵ^A also smoothly depending on the gauge orbit $\bar{\Gamma}$ to which $z \in \bar{\gamma}$ belongs. In this more general setting, $\epsilon^A(\tau|\bar{\Gamma}) \mapsto c^A(\tau|\bar{\Gamma}) \mapsto e^A(\tau|\bar{\Gamma})$, where $e^A(\tau|\bar{\Gamma})\bar{X}_A$ [and by identification $c^A(\tau|\bar{\Gamma})\bar{X}_A$] can be interpreted as the left-invariant Maurer–Cartan one-forms on $\mathcal{G}$ with values in the Lie algebra of the vertical vector fields on $\bar{\gamma}$ (or, by using the multitemporal equations in the Lie algebra of the left-invariant vector fields on G). If $e_A(\tau|\bar{\Gamma})$ are the left-invariant vector fields on $\mathcal{G}$ duals of $e_A(\tau|\bar{\Gamma})$ one has $i_{e_A(\tau|\bar{\Gamma})}(e^B(\tau|\bar{\Gamma})\bar{X}_B) = \bar{X}_A$. This is the Bonora–Cotta Ramusino[23] interpretation of ghosts, adapted to phase space. But now there is an interpretation also for the $\mathcal{P}_A$ (and therefore for the antighosts, which are present in the BRS Lagrangian as $\bar{c}_A \delta_s \psi^A$ and are proportional to $\mathcal{P}_A$ in the Kato–Ogawa approach). Since from Eqs. (65) we have

$$d\bar{f}(z(\tau)) = \frac{\partial \bar{f}}{\partial \tau} d\tau + \frac{\partial \bar{f}}{\partial \tau^A} d\tau^A = \frac{\partial \bar{f}}{\partial \tau} d\tau + (Y_A \bar{f})\theta^A$$
$$\doteq (\bar{X}_0 \bar{f}) d\tau + (\bar{X}_A \bar{f})\theta^A, \tag{81}$$

by writing $(Y_A \bar{f})\theta^A = -(Y_A \bar{f})\{c^A, \mathcal{P}_B\}\theta^B$ we get that $\mathcal{P}_A(\tau)\theta^A$ is the counterpart of $c^A(\tau) Y_A$. In the general case, this suggests that $\mathcal{P}_A(\tau|\bar{\Gamma})\theta^A$ are the left-invariant vector fields on $\mathcal{G}$ with values in the Maurer–Cartan one-forms on G. There is no counterpart of $c^A(\tau|\bar{\Gamma})\bar{X}_A \doteq c^A(\tau|\bar{\Gamma}) Y_A$, because $\bar{\gamma}$ is a presymplectic manifold so that the one-forms on $\bar{\gamma}$ dual of $\bar{X}_A$ are not uniquely defined. Equations (64) imply that on a one-parameter subgroup one has $\mathcal{P}_A \theta^A \mapsto \mathcal{P}_A \lambda^A(\tau) d\tau$. In the Kato–Ogawa approach we have a Hamiltonian $\bar{H} + \{\bar{\rho}, \bar{\Omega}\}$, $\bar{\Omega} = c^A \bar{\phi}_A + \cdots$, with $\bar{\rho}$ being the Hamiltonian gauge fixing corresponding to the Lagrangian one ψ^A. For consistency, when $\psi^A \doteq 0$, the extended Hamilton equations for a function $\bar{f}(q,p)$ must coincide with the Hamilton equations (61) but with the Dirac multipliers $\lambda^A(\tau)$ restricted to $\tilde{\lambda}^A(\tau)$ (the residual gauge freedom): therefore, $\mathcal{P}_A \theta^A \mapsto \mathcal{P}_A \tilde{\lambda}^A(\tau) d\tau \doteq \bar{\rho}\, d\tau$.

Let us make a last remark. Equations (71) imply

$$\delta_s^2 z^\alpha = -\{\delta_s z^\alpha, \bar{\Omega}\} = (\delta_s \bar{X}_A z^\alpha) c^A + (\bar{X}_A z^\alpha)\delta_s c^A$$
$$= \tfrac{1}{2}[([\bar{X}_B, \bar{X}_C] - C^A_{BC}\bar{X}_A) z^\alpha] c^B c^C = 0, \tag{82}$$

i.e., $\delta_s^2 = 0$ due to the phase space gauge algebra of the $\bar{X}_A$'s. Then similar conclusions have to hold at the Lagrangian level for the gauge algebra of the Lie–Bäcklund vector fields X_A associated to the $\bar{X}_A$ as shown in Ref. 1.

As was said in Sec. II, we could have chosen another arbitrary parametrization of the gauge transformations, i.e., another functional form $\bar{\phi}'_A$ for the constraints with structure functions $\bar{C}^A_{BC}(q,p)$. If for the corresponding δ_s we get $\delta_s^2 \neq 0$, this means that the vector fields $\bar{X}'_A$ or X'_A have some kind of pathology that forbids us to get a consistent global gauge algebra and then to use the Frobenius theorem to reconstruct the gauge part of the extremals.[1]

V. QUANTIZATION

Let us now suppose to quantize our theory with a parametrization of the gauge transformations such that globally one has $\{\bar{\phi}_A, \bar{\phi}_B\} = -C^C_{AB}\bar{\phi}_C$, i.e., such that $\delta_s^2 = 0$ classically. If we get the quantum algebra

$$[\hat{\phi}_A, \hat{\phi}_B] = C^C_{AB}\hat{\phi}_C, \quad [\hat{\phi}_A, \hat{H}] = 0, \tag{83}$$

the off-shell quantization[1] of the multitemporal equations (65) is

$$i\frac{\partial}{\partial \tau}\psi = \hat{H}\psi, \quad \psi = \psi(q|\tau, \tau^A),$$
$$iY_A \psi = \hat{\phi}_A \psi, \quad Y_A = B^B_A(\tau^C)\frac{\partial}{\partial \tau^B}, \tag{84}$$

and Eqs. (83) are the integrability conditions of Eqs. (84). Instead, the on-shell quantization of the Hamilton equations (69) for the observables is

$$i\frac{\partial}{\partial\tau}\Phi=\hat{H}\Phi,\qquad \Phi=\Phi(q^a|\tau,q^A)\begin{cases}a=1,\ldots,n-m,\\ A=1,\ldots,m.\end{cases}$$
$$\hat{\phi}_A\Phi=0, \tag{85}$$

This notation is oriented to constraints of the form $\hat{\phi}_A=\hat{p}_A-\hat{T}_A(\hat{q}^a,\hat{p}_a)$. Equations (85) are again integrable due to Eqs. (83). In Ref. 1 it is explained that the off-shell scalar product for the ψ (conserved with respect to τ and τ^A) has nothing to do with the physical on-shell one for the Φ (conserved with respect to τ and q^A), which are equivalence classes of the ψ's with respect to the unitary gauge transformations generated by the $\hat{\phi}_A$'s. The latter scalar product has a measure $d^{n-m}q^a$ and the q^A are now those coordinates q that are gauge degrees of freedom. Indeed, since we have done the quantization without the fixation of the gauge, the gauge freedom is still there in those configuration variables that carry a realization of the group manifold of G [in general a (possibly local) unitary transformation is needed to make this realization explicit]. A complete gauge fixing is obtained by restricting the solutions of Eqs. (85) to a "one-time" theory $q^A=\lambda^A(\tau)$ (choice of a one parameter subgroup in the group manifold of G); instead a partial gauge fixing may, for instance, be realized with the restriction $q^A=q^A(\rho^a)$, $\alpha=1,\ldots,k<m$ (choice of a k-parameters subgroup in the group manifold).

When there are ordering problems generating the central extension

$$[\hat{\phi}_A,\hat{\phi}_B]=C^C_{AB}\hat{\phi}_C+\hbar K_{AB},\quad [\hat{\phi}_A,\hat{H}]=0, \tag{86}$$

with $k_{AB}=-k_{BA}=$ const, we have a transition for a vector to a projective (or ray) unitary representation of the gauge transformations. This means[33] that if $U(\tau^A_1),U(\tau^A_2)$ are the unitary operators corresponding to V_1, $V_2\in G$, respectively (with τ^A_1,τ^A_2 being their group manifold coordinates) and $\rho^A(\tau_1,\tau_2)$ are the coordinates of V_1V_2 from the group composition law, then $U(\tau_1)U(\tau_2)=e^{i\Phi(\tau_1,\tau_2)}U(\rho(\tau_1,\tau_2))$ and $U^{-1}(\tau_1)=e^{-i\Phi(\tau_1,\tau_1)}U(\bar{\tau}_1)$, where $\bar{\tau}^A_1$ are the coordinates of V^{-1}. Here Φ is called an exponent and it is a two-cocycle. By defining $\varphi_A(\tau)=\partial\Phi(\tau',\tau)/\partial\tau'^A|_{\tau^A=0}$, called the "right generators," one gets[33]

$$Y_A\varphi_B-Y_B\varphi_A=C^C_{AB}\varphi_C+t_{AB}, \tag{87}$$

where the constants $t_{AB}=-t_{BA}$ are the said ray constants of the projective representation and satisfy $C^D_{AB}t_{CD}+C^D_{CA}t_{BD}+C^D_{BC}t_{AD}=0$.

Therefore if the $\hbar k_{AB}$'s in Eqs. (86) satisfy these conditions, the on-shell quantization equations (85) are replaced by the following ones:

$$i\frac{\partial}{\partial\tau}\Phi=\hat{H}\Phi,\quad \Phi=\Phi(q^a|\tau,q^A),$$
$$\hat{\phi}_A\Phi=\varphi_A\Phi,\quad \varphi_A=\varphi_A(q), \tag{88}$$

with the functions $\varphi_A(q)$ solutions of the equations

$$\hat{\phi}_A\varphi_B-\hat{\phi}_B\varphi_A=C^C_{AB}\varphi_C+\hbar K_{AB}, \tag{89}$$

because Eqs. (89) and (86) are the integrability conditions of Eqs. (88).

With more severe ordering problems, the canonical quantization is inconsistent (anomalies in the form of Schwinger terms). The same procedure has to be followed with every parametrization having $\delta^2_s=0$: in general, it will be possible to have a consistent canonical quantization only of a subset of these parametrizations.

The quantum BRS method starts with the quantization of the coordinates $q^i, p_i, c^A, \mathscr{P}_A$ of the extended phase space (we are using the Kato–Ogawa approach but the same holds with the BFV one): $[\hat{q}^i,\hat{p}_j]=i\hbar\delta^i_j$, $[\hat{c}^A,\hat{\mathscr{P}}_B]_+=-i\delta^A_B$, $\hat{c}^{A\dagger}=\hat{c}^A$, $\hat{\mathscr{P}}^\dagger_A=-\hat{\mathscr{P}}_A$. If the ordering problems for the chosen parametrization (with $\delta^2_s=0$) of the constraints $\bar{\phi}_A$ allow us to build a Hermitian BRS operator $\hat{\Omega}=\hat{c}^A(\hat{\phi}_A-(1/2)C^C_{AB}\hat{c}^B\hat{\mathscr{P}}_C-(i/2)C^B_{AB})$ such that $\{\bar{\Omega},\bar{\Omega}\}=0$ goes into $[\hat{\Omega},\hat{\Omega}]_+=2\hat{\Omega}^2=0$ we get a consistent BRS quantization (maybe in some critical dimension), which is compatible with an indefinite metric ($\bar{\Omega}|\rangle$ is a zero norm state). Since the original gauge freedom has been reduced to a residual one with the BRS gauge-fixing method, the BRS quantization lies in between the off- and the on-shell quantizations. Instead of Eqs. (84) and (85), we get the following Schrödinger equation with the BRS Hamiltonian:

$$i\frac{\partial}{\partial\tau}\Psi=\left(\hat{H}+[\hat{\rho},\hat{\Omega}]_+\right)\Psi \tag{90}$$

and the BRS states $\Psi_{\rm BRS}$ are selected by the requirement $\hat{\Omega}\Psi_{\rm BRS}=0$ (so that $\Psi_{\rm BRS}\in\operatorname{Ker}\hat{\Omega}$), $\hat{N}_g\Psi_{\rm BRS}=0$, where $\hat{N}_g=(i/2)(\hat{c}^A\hat{\mathscr{P}}_A-\hat{\mathscr{P}}_A\hat{c}^A)=-\hat{N}^\dagger_g=i\hat{c}^A\hat{\mathscr{P}}_A-1/2$ is the ghost number operator and $[\hat{\Omega},\hat{N}_g]=\hat{\Omega}$. Again these BRS states are defined modulo the trivial zero norm BRS states $\hat{\Omega}\Psi\in\operatorname{Im}\hat{\Omega}$ (the residual gauge freedom) and the final physical states are the equivalence classes with respect to these zero norm states. This is called the quantum BRS cohomology,[29,34] $\operatorname{Ker}\hat{\Omega}/\operatorname{Im}\hat{\Omega}$, which does not contain zero norm states if $\hat{\Omega}$ is "complete"[34]: i.e., if $\operatorname{Im}\hat{\Omega}=\operatorname{Ker}_0\hat{\Omega}$, the set of all the zero norm states in $\operatorname{Ker}\hat{\Omega}$; then $\operatorname{Ker}\hat{\Omega}$ does not contain negative norm states.

This quantization procedure gives rise to interpretational problems[35] about the necessity of the condition $\hat{N}_g\Psi_{\rm BRS}=0$ (inequivalent quantum theories could be generated by relaxing it), about the vacuum states for $\hat{c}^A,\hat{\mathscr{P}}_A$, and about the finiteness of the scalar product [so that the use of an off-shell-like scalar product produces a Gupta–Bleuler quantization of 1st-class constraints: also, in the multitemporal approach, the solutions of Eqs. (85) have an infinite norm in the off-shell scalar product and this is cured by realizing that these solutions are equivalence classes of the off-shell solutions and therefore require a new on-shell scalar product].

Our attitude is that when $\hat{\Omega}^2=0$ one is realizing a partial gauge fixing of the unitary gauge transformations generated by the $\hat{\phi}_A$'s, which satisfy either Eq. (83) or Eq. (86), and that the physical theory should correspond either to Eqs. (85) or to Eqs. (88), with the corresponding partial gauge fixing of the parameters q^A.

Equation (90) should be implemented with a scalar product $\langle,\rangle$ constant in τ. Moreover, if Ψ is a solution of Eq. (90), $\hat{\Omega}\Psi$ is also. When one restricts Ψ to be a BRS state $\Psi_{\rm BRS}$, Eq. (90) becomes

$$i\frac{\partial}{\partial\tau}\Psi_{\rm BRS}=\Big(\hat H+\hat\Omega\hat\rho\Big)\Psi_{\rm BRS},\quad \hat\Omega\Psi_{\rm BRS}=0, \tag{91}$$

where $\hat\Omega\hat\rho$ describes the residual gauge transformations. Since, in general, $\Psi_{\rm BRS}$ is not normalizable by using $\langle,\rangle$, a new scalar product, $\langle,\rangle_{\rm BRS}$ has to be found with the same methodology like for Eqs. (85). The original variables in the Ψ of Eq. (90) have to be divided into two groups dictated by the way in which the equation $\hat\Omega\Psi_{\rm BRS}=0$ is solved. One of these groups of variables will form the parameters of the residual gauge transformations. The scalar product $\langle,\rangle_{\rm BRS}$ will only contain an integration on the other group of variables and will have to be conserved with respect to the gauge parameters beside with respect to τ. To find $\langle,\rangle_{\rm BRS}$ will be the most difficult part of the BRS quantization like it happens with Eqs. (85).[5]

What makes things complicated is the fact that in the BRS quantization, one is quantizing also the group manifold and the ghost-dependent part of the solutions of Eq. (90) describes the quantum analog of the determination of the residual gauge freedom, which at the classical level was obtained by means of $\rho c^A(\tau)\doteq\bar c^A(\tau)$. Since the ghosts and their momenta go into the Faddeev–Popov ghosts only in a limit in which the gauge is completely fixed (see the Faddeev–Popov measure), we get a nontrivial quantum mechanics for the ghosts, which must reduce to the trivial one for the Faddeev–Popov ghosts when the gauge is fixed. Even if this nontrivial quantum mechanics seems to be the price for a quantum control of the hypothesis of gauge algebra, it is not clear to us the meaning of things like the doubling of the vacuum states.

Let us remark that the selection of a representative Ψ_F (a true physical state) from each class ($\Psi_{\rm BRS}+\hat\Omega\Psi$) can be done, as in the classical case, by requiring that it satisfies the quantum equivariance condition. Equation (73) and (74) become

$$\begin{aligned}\hat H_A=\hat H_A^\dagger&=-\tfrac12 C_{AB}^{C}(\hat c^B\hat{\mathscr P}_C-\hat{\mathscr P}_C\hat C^B)\\&=-C_{AB}^{C}\hat c^B\hat{\mathscr P}_C-\tfrac{i}{2}C_{AB}^{B},\end{aligned}\tag{92}$$

$$\begin{aligned}&[\hat\phi_A+\hat H_A,\hat\phi_B+\hat H_B]=-iC_{AB}^{C}(\hat\phi_C+\hat H_C),\\&\hat\Omega=\hat c^A(\hat\phi_A+\hat H_A+\tfrac12 C_{AB}^{c}\hat c^B\hat{\mathscr P}_C),\end{aligned}\tag{93}$$

and we get for an equivariant wave function $\Psi_{\rm BRS}^E$

$$\begin{aligned}&(\hat\phi_A+\hat H_A)\Psi_{\rm BRS}^E=0,\\&0=\hat\Omega\Psi_{\rm BRS}^E\\&\quad=\tfrac12 C_{AB}^{C}\hat c^A\hat c^B\hat{\mathscr P}_C\Psi_{\rm BRS}^E\\&\quad=-\tfrac12\hat c^A\Big(\hat H_A+\frac{i}{2}C_{AB}^{B}\Big)\Psi_{\rm BRS}^E\\&\quad=\tfrac12\hat c^A(\hat\phi_A-(i/2)C_{AB}^{B})\Psi_{\rm BRS}^E\\&\Rightarrow\hat c^A\Psi_{\rm BRS}^E=0\\&\Rightarrow\hat\phi_A\Psi_{\rm BRS}^E=-\hat H_A\Psi_{\rm BRS}^E\\&\quad=\tfrac12 C_{AB}^{C}\hat c^B\hat{\mathscr P}_C\Psi_{\rm BRS}^E=-\frac{i}{2}C_{AB}^{B}\Psi_{\rm BRS}^E.\end{aligned}\tag{94}$$

This last result is contradictory because $\hat\phi_A=\hat\phi_A^\dagger$ so that it must also be $\hat{\mathscr P}_A\Psi_{\rm BRS}^E=0$. Since the wave function Ψ belongs to the tensor product $H_E\otimes H_O$ of an even and an odd Hilbert space ($\Psi=\Psi_E\otimes\Psi_O$), the equations $\hat c^A\Psi_{\rm BRS}^E=\hat{\mathscr P}_A\Psi_{\rm BRS}^E=0$ imply that $\Psi_{\rm BRS}^E=\Psi_F\otimes\{\Psi_O=0\}$, where $\{\Psi_O=0\}$ is the origin of the vector space H_O (this same mechanism works for the Gupta–Bleuler method applied to even 2nd-class constraints[1]). Then, Eqs. (94) imply

$$\hat\phi_A\Psi_F=0. \tag{95}$$

Again, the equivariance condition eliminates the residual gauge freedom and reproduces the on-shell quantum theory.

As a final remark let us note that we are very far from understanding how to demonstrate the physical equivalence of the quantum theories coming from different BRS quantizations of the same system, due to the completely formal nature of the Fradkin–Vilkovisky theorem. *A priori* we could get nonunitarily equivalent Hilbert spaces, different critical dimensions and so on: What should we choose for the good quantum theory?

VI. CONCLUSION

We hope to have succeeded in presenting a unified picture of the various aspects of the BRS approach, at least at a qualitative level. Each point of view shows some of the features of the theory, but often the differences in the technical langauges produce difficulties in the comparison of the results and may obscure the concepts. Another problem is that usually a physicist has only a local knowledge of the system and is only able to define the action of some local Lie algebra on it. Instead, most of the beautiful mathematical results he would like to use assume that a global action of a Lie group is given. Once more the necessity of a global description of the physical system from the very beginning turns out to be one of the most relevant problems. Actually, one often tries to adapt methods, having their justification only in the well-defined context of Lie algebras and groups, to situations in which they can at best work locally, so that one looses any global control. For instance, this is the case when one has structure functions instead of structure constants in the Poisson algebra of the constraints. Another example is the problem to find which is the really suited cohomology theory to be associated with the BRS method. Part of the problem is that one starts with a Lagrangian description and the cohomological description should be adapted to it.[31] Instead, it turns out easier to study the cohomological aspects of the associated Hamiltonian description, which in any case has its own problems, especially in field theory. However, also here there is the hard problem that the Legendre transformation of a singular Lagrangian identifies only the primary constraint manifold: this is a presymplectic manifold (there is no concept of Poisson brackets), which is coisotropically embedded[36] in the reference phase space in which the Dirac–Bergmann theory is defined. One should try to adapt the concepts of Lagrangian gauge fixing and BRS invariance to presymplectic manifolds and then look for a cohomology theory adapted to them (see Ref. 29 for attempts in this direction). As a preliminary step one has to develop the differential geometry of exact presymplectic manifolds, the ones associated with singular Lagrangians, which has not yet been investigated. This will be the subject of a future paper, which will be the natural continuation of the study of the theory of singular Lagrangians and Hamiltonian constraints, initiated with Refs. 1 and 2 and prosecuted with the present paper.

[1]L. Lusanna, CERN preprint TH.5087/88 (1988), (to appear in Phys. Rep.).

[2]L. Lusanna, "The Second Nœther Theorem as the Basis of the Theory of Singular Lagrangians and Hamiltonian Constraints," Firenze University preprint (1989) (Submitted to Lecture Notes in Physics).

[3]C. Becchi, A. Rouet, and R. Stora, Phys. Lett. B **52**, 344 (1974); CNRS Centre de Physique Thèorique, Marseille, Colloquium on Recent Progress in Lagrangian Field Theory, June 1974; Commun. Math. Phys. **42**, 127 (1975); Ann. Phys. (N.Y.) **98**, 287 (1976).

[4]W. Sarlet and F. Cantrijn, SIAM Rev. **23**, 467 (1981).

[5]G. Longhi and L. Lusanna, Phys. Rev. D **34**, 3707 (1986).

[6]R. Stora, "Differential Algebras in Field Theory," Talk at the Second Summer Meeting on Quantum Mechanics of Fundamental Systems, Santiago, Chile, 1987.

[7]A. Neveu and P. West, Phys. Lett. B **182**, 343 (1986).

[8]K. Kato and K. Ogawa, Nucl. Phys. B **212**, 443 (1983).

[9]E. S. Fradkin and G. A. Vilkovisky, Phys. Lett. B **55**, 224 (1975); preprint CERN TH.2332 (1977) (unpublished); I. A. Batalin and G. A. Vilkovisky, Phys. Lett. B **69**, 309 (1977).

[10]M. Henneaux, Phys. Rep. **126**, 1 (1985); *Classical Foundations of BRS Symmetry* (Bibliopolis, Naples Italy, 1988).

[11]J. Govaerts, preprint CERN TH.5010/88 (1988).

[12]T. Kugo and I. Ojima, Suppl. Prog. Theor. Phys. **66**, 1 (1979).

[13]G. Parisi and N. Sourlas, Phys. Rev. Lett. **43**, 744 (1979).

[14]L. Baulieu, Phys. Rep. **129**, 1 (1985).

[15]W. Siegel and B. Zweibach, Nucl. Phys. B **282**, 125 (1987). W. Siegel, Nucl. Phys. B **284**, 632 (1987).

[16]H. Aratyn, R. Ingermanson, and A. J. Niemi, Phys. Lett. B **189**, 427 (1987).

[17]A. Barducci, R. Casalbuoni, D. Dominici, and R. Gatto, Phys. Lett. B **187**, 135 (1987); **194**, 257 (1987); Genève Univ. preprint UGVA-DPT 1988/11-598. R. Casalbuoni, D. Dominici, R. Gatto, and J. Gomis, Phys. Lett. B **198**, 177 (1987).

[18]C. Itzykson and B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1985).

[19]R. Fanelli, J. Math. Phys. **17**, 490 (1976).

[20]M. Henneaux and C. Teitelboim, Ann. Phys. (NY) **143**, 127 (1982).

[21]B. Konstant and S. Sternberg, Ann. Phys. (NY) **176**, 49 (1987).

[22]R. Loll, Commun. Math. Phys. **119**, 509 (1988).

[23]L. Bonora and P. Cotta Ramusino, Commun. Math. Phys. **87**, 589 (1983).

[24]L. P. Eisenhart, *Continuous Groups of Transformations* (Princeton U.P., Princeton, NJ, 1933). R. O. Fulp and J. A. Marlin, Pacific J. Math. **67**, 373 (1976); Rep. Math. Phys. **18**, 295 (1980). G. Marmo, in *Proceedings of the Symposium on Modern Developments in Analytical Mechanics*, Torino 1982, Atti della Accademia delle Scienze di Torino, Suppl. Vol. 117 (1983).

[25]A. Trautman, *Differential Geometry for Physicists* (Bibliopolis, Naples, Italy, 1984).

[26]G. Marmo, E. J. Saletan, A. Simoni, and B. Vitale, *Dynamical Systems* (Wiley, Chichester, 1985).

[27]L. Lusanna, in *Proceedings IV M. Grossmann Meeting*, Rome 1985, edited by R. Ruffini (Elsevier, Amsterdam, 1986); G. Longhi, L. Lusanna, and J. M. Pons, J. Math Phys. **30**, 1893 (1989).

[28]E. C. G. Sudarshan and N. Mukunda, *Classical Mechanics: A Modern Perspective* (Wiley, New York, 1974).

[29]M. Henneaux and C. Teitelboim, Commun. Math. Phys. **115**, 213 (1988).

[30]D. McMullan, Imperial College preprint TP/83-84/21 (1984) (unpublished); Phys. Rev. D **33**, 2501 (1986); J. Math. Phys. **28**, 428 and 438 (1987). D. McMullan and J. Paterson, Univ. of Glasgow preprint GUPTA/87/12-1 (1987).

[31]R. Schmidt, Emory Univ. preprint, Atlanta (1988).

[32]P. Cotta Ramusino and C. Reina, J. Geom. Phys. **1**, 121 (1984).

[33]V. Bargmann, Ann. Math. **59**, 1 (1954). J. Krause, J. Math. Phys. **28**, 2268 (1987).

[34]M. Spiegelglas, Nucl. Phys. B **283**, 205 (1987). J. M. Figueroa-O'Farrill and T. Kimura, SUNY Stony Brook preprint ITP-SB-88-34 (1988).

[35]S. Hwang and R. Marnelius, Göteborg Univ. preprint 88-35 (UTTG-14-18) (1988). R. Marnelius, "Generalized BRS Quantization," Talk presented at the Conference Geometrical and Algebraic Aspects of Nonlinear Field Theory, Naples-Amalfi, 1988; Nucl. Phys. B **294**, 671 and 685 (1987); Phys. Lett. B **99**, 467 (1981).

[36]M. J. Gotay and J. Sniatycki, Commun. Math. Phys. **82**, 377 (1981); M. J. Gotay, Proc. Am. Math. Soc. **84**, 111 (1982).

Spinor solution in a Kasner universe

K. D. Krori, P. Borgohain, P. Kar, and Dipali Das (Kar)
Mathematical Physics Forum, Cotton College, Gauhati-781001, India

(Received 12 August 1988; accepted for publication 6 September 1989)

In this paper an exact solution of the Dirac equation in an anisotropic Kasner metric is presented.

I. INTRODUCTION

It is now almost established that the quantum-mechanical effects play a crucial role in the evolution of space-time near singularity, as in the last stages of a collapsing star or in the early stages of the universe, viz., a few seconds after the big bang. This is why recently much interest has been evinced by a large number of workers to study quantum theory in curved space-time and to examine the creation of particles and antiparticles in curved space-time.[1-4] Recently Chimento and Mollerach[5,6] have studied the particle creation in Robertson–Walker metrics and extended the study to Bianchi type I metrics. In this paper we examine the Dirac equation in the curved space-time represented by an anisotropic Kasner metric. The exact solution of the equation is derived.

II. GENERALIZED DIRAC EQUATION AND SOLUTION

The generalized Dirac equation in curved space-time is

$$[\gamma^\mu \nabla_\mu - m]\psi(x,t) = 0, \tag{1}$$

where

$$\nabla_\mu = \partial_\mu - \sigma_\mu \tag{2}$$

and

$$\sigma_\mu = \tfrac{1}{4}\gamma^{(\alpha)}\gamma^{(\beta)} V_\alpha{}^\nu V_{(\beta)\nu,\mu}. \tag{3}$$

Here σ_μ are spinorial affine connections, $\gamma^{(\alpha)},\gamma^{(\beta)}$ are flat space-time Dirac matrices, γ^μ are curved space-time Dirac matrices, $V_{(\alpha)}{}^\nu$ are four-vector fields called vierbein and are related to the metrics by the equation

$$V_{(\alpha)}{}^a V_{(\beta)}{}^b \eta^{(\alpha)(\beta)} = g^{ab}; \tag{4}$$

$\eta^{(\alpha)(\beta)}$ are Minkowski metric with signature

$$\{-1,\ -1,\ -1,\ +1\}.$$

Also, the Dirac matrices in the two space-times are connected by the relation

$$\gamma^\mu = V_{(\alpha)}{}^\mu \gamma^{(\alpha)}. \tag{5}$$

Let us now take the anisotropic Kasner metric in the form

$$ds^2 = dt^2 - t^{4/3}\,dx_1^2 - t^{4/3}\,dx_2^2 - t^{-2/3}\,dx_3^2. \tag{6}$$

With the help of Eq. (4), vierbeins worked out for the metric (6) are

$$V_{(1)}{}^1 = t^{-2/3} = V_{(2)}{}^2,\quad V_{(3)}{}^3 = t^{1/3},\quad V_{(0)}{}^0 = 1. \tag{7}$$

Now with the help of (5) and (7) the Dirac matrices in curved space-time may be obtained as

$$\gamma^1 = t^{-2/3}\gamma^{(1)},\quad \gamma^2 = t^{-2/3}\gamma^{(2)},\quad \gamma^3 = t^{1/3}\gamma^{(3)},\quad \gamma^0 = \gamma^{(0)}. \tag{8}$$

In Eq. (3) $V_{(\beta)\nu,\mu}$ are expressed as

$$V_{(\beta)\nu,\mu} = \partial_\mu V_{(\beta)\nu} - \Gamma_{\nu\mu}{}^\lambda V_{(\beta)\lambda}, \tag{9}$$

where $\Gamma_{\nu\mu}{}^\lambda$ are Christoffel's symbols whose nonvanishing components are

$$\begin{aligned}&\Gamma^0_{11} = \tfrac{2}{3}t^{1/3} = \Gamma^0_{22},\\ &\Gamma^1_{10} = \Gamma^1_{01} = 2/3t = \Gamma^2_{20} = \Gamma^2_{02},\\ &\Gamma^0_{33} = -(1/3t^{5/3}),\\ &\Gamma^3_{30} = -1/3t = \Gamma^3_{03}.\end{aligned} \tag{10}$$

The connections σ_μ can now be written as

$$\begin{aligned}&\sigma_1 = (1/3t^{1/3})\gamma^{(0)}\gamma^{(1)},\\ &\sigma_2 = (1/3t^{1/3})\gamma^{(0)}\gamma^{(2)},\\ &\sigma_3 = -(1/6t^{4/3})\gamma^{(0)}\gamma^{(3)},\\ &\sigma_0 = 0.\end{aligned} \tag{11}$$

The Dirac equation (1) can now be solved by the method of separation of variables. We take the solution in the form

$$\psi(x,t) = (8\pi^3 t)^{-1/2}\begin{pmatrix} e^{i\int_0^t \eta_p\,dt} \\ e^{i\int_0^t \xi_p\,dt} \\ e^{i\int_0^t A_p\,dt} \\ e^{i\int_0^t B_p\,dt}\end{pmatrix} e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{12}$$

Inserting (12) in (1) we get the following matrix equation:

$$\begin{bmatrix} (\eta_p - m) & 0 & \bar{E}_3 & E_2 \\ 0 & (\xi_p - m) & E_1 & -\bar{E}_3 \\ \bar{E}_3 & E_2 & (A_p + m) & 0 \\ E_1 & -\bar{E}_3 & 0 & (B_p + m)\end{bmatrix} \times \begin{bmatrix} e^{i\int_0^t \eta_p\,dt} \\ e^{i\int_0^t \xi_p\,dt} \\ e^{i\int_0^t A_p\,dt} \\ e^{i\int_0^t B_p\,dt}\end{bmatrix} = 0, \tag{13}$$

where

$$\begin{aligned}&E_1 = (k_1/t^{2/3}) + (i(k_2/t^{2/3})) = \bar{E}_1 + i\bar{E}_2,\\ &E_2 = (k_1/t^{2/3}) - (i(k_2/t^{2/3})) = \bar{E}_1 - i\bar{E}_2,\\ &E_3 = k_3 t^{1/3},\end{aligned} \tag{14}$$

 0022-2488/90/020446-02$03.00

and

$$\mathbf{k}\cdot\mathbf{x} = k_1x_1 + k_2x_2 + k_3x_3.$$

Equation (13) will have nontrivial solutions if

$$[E^2 - (B_p + m)(\eta_p - m)][E^2 - (A_p + m)(\xi_p - m)] + \overline{E}_3^2(\eta_p - \xi_p)(B_p - A_p) = 0,$$

with

$$E^2 = \overline{E}_1^2 + \overline{E}_2^2 + \overline{E}_3^2. \quad (15)$$

For simplicity we now consider that the particle travels along the z direction so that $k_3 = k$ and $k_1 = 0 = k_2$. Then Eq. (13) reduces to

$$\begin{bmatrix} (\eta_p - m) & 0 & kt^{1/3} & 0 \\ 0 & (\xi_p - m) & 0 & -kt^{1/3} \\ kt^{1/3} & 0 & (A_p + m) & 0 \\ 0 & -kt^{1/3} & 0 & (B_p + m) \end{bmatrix} \times \begin{bmatrix} e^{i\int_0^t \eta_p\,dt} \\ e^{i\int_0^t \xi_p\,dt} \\ e^{i\int_0^t A_p\,dt} \\ e^{i\int_0^t B_p\,dt} \end{bmatrix} = 0. \quad (16)$$

Solving Eq. (16) and introducing the condition that

$$(\eta_p - m)(A_p + m) = k^2t^{2/3} = (B_p + m)(\xi_p - m), \quad (17)$$

the column matrix may now be expressed as

$$\begin{bmatrix} e^{i\int_0^t \eta_p\,dt} \\ e^{i\int_0^t \xi_p\,dt} \\ e^{i\int_0^t A_p\,dt} \\ e^{i\int_0^t B_p\,dt} \end{bmatrix} = \begin{bmatrix} (A_p + m) - kt^{1/3} \\ (B_p + m) - kt^{1/3} \\ (\eta_p - m) - kt^{1/3} \\ kt^{1/3} - (\xi_p - m) \end{bmatrix}. \quad (18)$$

Operating Eq. (16) from the left with the operator $[\gamma^{(0)}\partial_0 + ikt^{1/3}\gamma^{(3)} - m]$ and substituting (18) we get the following four differential equations in

η_p, ξ_p, A_p, and B_p:

$$\dot{\eta}_p[A_p - kt^{1/3} + m] + i\eta_p[kt^{1/3}A_p - mA_p + mkt^{1/3} - m^2 - (i/3)kt^{-2/3}] + [-im(kt^{1/3}A_p - mA_p + mkt^{1/3} - m^2 - (i/3)kt^{-2/3}) + ik^2(-kt + mt^{2/3} + (i/3)t^{-1/3})] = 0, \quad (19)$$

$$\dot{\xi}_p[B_p - kt^{1/3} + m] + i\xi_p[kt^{1/3}B_p - mB_p + mkt^{1/3} - m^2 - (i/3)kt^{-2/3}] + [-im(kt^{1/3}B_p - mB_p + mkt^{1/3} - m^2 - (i/3)kt^{-2/3}) + ik^2(-kt + mt^{2/3} + (i/3)t^{-1/3})] = 0, \quad (20)$$

$$\dot{A}_p[\eta_p - kt^{1/3} - m] + iA_p[kt^{1/3}\eta_p + m\eta_p - mkt^{1/3} - m^2 - (i/3)kt^{-2/3}] + [im(kt^{1/3}\eta_p + m\eta_p - mkt^{1/3} - m^2 - (i/3)kt^{-2/3}) + ik^2(-kt - mt^{2/3} + (i/3)t^{-1/3})] = 0, \quad (21)$$

$$\dot{B}_p[\xi_p - kt^{1/3} - m] + iB_p[kt^{1/3}\xi_p + m\xi_p - mkt^{1/3} - m^2 - (i/3)kt^{-2/3}] + [im(kt^{1/3}\xi_p + m\xi_p - mkt^{1/3} - m^2 - (i/3)kt^{-2/3}) + ik^2(-kt - mt^{2/3} + (i/3)t^{-1/3})] = 0. \quad (22)$$

All four of the above differential equations are similar to the Abel equation of the second kind and may be solved accordingly. With suitable values of integration constants, the solutions are

$$\begin{aligned} \eta_p &= -kt^{1/3} + m, \\ \xi_p &= -kt^{1/3} + m, \\ A_p &= -kt^{1/3} - m, \\ B_p &= -kt^{1/3} - m. \end{aligned} \quad (23)$$

We thus obtain the exact solution of the Dirac equation in a Kasner metric.

ACKNOWLEDGMENTS

The authors express their profound gratitude to the government of Assam for all facilities provided at Cotton College, Gauhati, India to carry out the investigations reported in this paper.

[1] L. Parker, Phys. Rev. **183**, 1057 (1969).
[2] L. Parker, Phys. Rev. D **3**, 346 (1971).
[3] B. De Witt, Phys. Rep. **19**, 297 (1975).
[4] M. Castagnino, L. Chimento, D. Harari, and C. Nunez, J. Math. Phys. **25**, 360 (1984).
[5] S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).
[6] L. Chimento and M. S. Mollerach, Phys. Lett. A **121**, 7 (1987).

Gauge transformations in Dirac theory of constrained systems

Carlos A.P. Galvão
Centro Brasileiro de Pesquisas Físicas—CBPF/CNPq, Rua Dr. Xavier Sigaud, 150, 22290 Rio de Janeiro, Brazil
João Batista T. Boechat
DEPEX–PETROBRÁS, Rio de Janeiro, Brazil

(Received 30 May 1989; accepted for publication 13 September 1989)

According to Dirac's prescription the generator of gauge transformations for a constrained system endowed with primary and secondary first class constraints is constructed as a linear combination of all these (first class) constraints. Using the total Hamiltonian to generate the dynamics of the system it is shown that the time evolution of the coefficients of the secondary constraints in the generator of gauge transformations is not independent but is determined by the coefficients of the primary constraints. This result is applied to some physically interesting systems.

I. INTRODUCTION

Since Dirac[1] developed his theory for constrained systems there has been considerable progress in the understanding of those systems. The interest on his theory is indeed justified, not only for the deep insight it provides into the conceptual framework but also for the very powerful techniques it provides, which can be applied to a very broad class of important physical systems.

One point in Dirac's theory that has been the target of criticism by some authors[2–4] is his conjecture that all first class constraints are generators of gauge transformations. He also introduced the concept of an extended Hamiltonian that includes all the first class constraints and generates the dynamical evolution of the system with full gauge freedom. In spite of the lack of a proof of his conjecture (or even a proof that it is not correct) we do not know of any physically important system to which Dirac's conjecture leads to the *wrong* result.

In order to obtain all the constraints of a theory, one must use Dirac's algorithm, which in some cases is very tedious. But once all the first class constraints are obtained one can construct a generator of gauge transformations as a linear combination of those constraints, the coefficients of which are, in principle, arbitrary. Application of this procedure[5] to the case of Yang–Mills theory requires a by-hand adjustment of the coefficients in order to match the result with the well known gauge transformation law for the Yang–Mills potentials in the Lagrangian form.

The example of Yang–Mills theory suggests questioning the degree of arbitrariness of the coefficients that appear in the generator. Admitting that the evolution of a given dynamical system is generated by the total Hamiltonian H_T (we remark that this poses no restriction on the dynamics) we compared two trajectories of the system corresponding to the same initial data but to different choices of the arbitrary functions in H_T. Taking into account that the physical states of the system cannot depend on the choice of the arbitrary functions, the corresponding states along the two trajectories must be related by a gauge transformation. The result of the procedure is a differential equation relating the coefficients of the primary and secondary first class constraints. The generator so obtained has been applied to various sytems yielding the correct results. (For different approaches to obtaining the generator of gauge transformations, see Refs. 6 and 7.)

The paper is organized as follows. In Sec. II, we discuss some aspects of Dirac's theory relevant for the following sections. In Sec. III, we present our approach for obtaining the generator. Section IV is devoted to applications.

II. THE GENERATOR OF GAUGE TRANSFORMATIONS ACCORDING TO DIRAC'S THEORY

Let us consider the evolution of a mechanical system in phase space with canonical coordinates (q^n,p_n), $n = 1,...,N$. We suppose that the system is singular and denote the full set of independent constraints (to be specified later on) by $\bar{\phi}_i \approx 0$, $i = 1,...,m$, which define a subspace $\mathscr{M}$ in phase space, where the motion of the system actually occurs.

According to Dirac's theory the total Hamiltonian for the system is defined as

$$H_T = H_c + u^k(q,p)\phi_k, \qquad (2.1)$$

where H_c is the canonical Hamiltonian ϕ_k, $k = 1,...,K$, are the primary constraints, and $u^k(q,p)$ are arbitrary functions. The constraints ϕ_k constitute a subset of the constraints C_i. In principle, the primary constraints are known once the momenta are calculated and are incorporated in the Hamiltonian by the method of Lagrange multiplers.

The consistency conditions of time preservation of the primary constraints, $\dot{\phi}_k = \{\phi_k,H_T\} \approx 0$, in general lead to the existence of new constraints, ψ_l, which are called secondary constraints. (During this process some of the functions u^k can possibly be determined, but whether this happens or not is not important in what follows.) The set of constraints C_i is then constituted by all the primary and secondary constraints. For simplicity, we will suppose that this set is first class. The important property of this set of constraints is that, together with H_c, it constitutes an algebra (denoted by $\mathscr{G}$) under the Poisson bracket operation. Indeed, one can easily show that, for arbitrary linear combinations g_i of elements of $\mathscr{G}$, with arbitrary coefficients depending on (q^k,p_k), the following relations hold:

 0022-2488/90/020448-04$03.00

$$\{g_i,g_j\} = C^k{}_{ij}(q,p)g_k,$$
$$\{g_i,H_c\} = C^k{}_i(q,p)g_k. \tag{2.2}$$

(Instead of H_c we should use $H_0 = H_c + \lambda_n\phi_n$, where λ_n are the multipliers determined during the consistency procedure. But as we said before this is not important for our purpose.) It follows that the set $(\bar{\phi}^i, H_c)$ constitutes a basis in $\mathscr{G}$. As generators of infinitesimal transformations the elements of $\mathscr{G}$ map $\mathscr{M}$ on $\mathscr{M}$. When the coefficients in (2.2) are constants, $\mathscr{G}$ is a Lie algebra to which is associated the group of infinitesimal transformations on $\mathscr{M}$.

Now, given the initial data $(q^k,p_k)_{t=t_0}$, the physical state of the system is well determined at t_0. However, the time evolution of the system generated by the total Hamiltonian leads to the appearance of the arbitrary functions in the solutions of the equations of motion. This implies that at later times there are several physically equivalent sets of canonical variables that evolve from the same initial data. In other words, for each choice of the arbitrary functions u^k there is an extremal curve or trajectory of the system, starting at $(q^k,p_k)_{t=t_0}$.

From the physical point of view the choice of the arbitrary functions is irrelevant in the sense that the corresponding states of the system must be equivalent. Hence one is led to say that the terms involving the primary first class constraints in H_T generate transformations that do not change the physical states of the system. In order words, they generate gauge transformations.

What is clear from the above discussion is that not only are the primary first class constraints generators of gauge transformations but also the secondary (first class) ones; Dirac conjectured that they should also be included in the Hamiltonian and defined the extended Hamiltonian

$$H_E = H_T + v^i(q,p)\bar{\phi}_i, \tag{2.3}$$

which generates the evolution of the system with full gauge freedom. In spite of their completely different physical origins, it is perfectly acceptable from the physical point of view that all the first class constraints must be treated on equal footing.

According to Dirac's prescription the generator of gauge transformations for the system can be written as

$$G = \Omega^i(q,p)\bar{\phi}_i, \tag{2.4}$$

where $\Omega^i(q,p)$ are arbitrary functions and $(\bar{\phi}_i)$ denotes all the first class constraints. A straightforward application of (2.4) to the important case of Yang–Mills theory requires an adjustment of the "arbitrary" functions at the final step in order to recover the correct Lagrangian transformation law for the gauge potentials, namely $\delta A^\mu_a = D^\mu\omega_a(x)$. [Another procedure[8] makes use of the equations of motion generated by H_e so as to eliminate some of the arbitrary functions and to identify the remaining ones with $A^0_a(x)$.] Guided by these facts we questioned the degree of arbitrariness of the functions $\Omega_i(q,p)$ that appear in (2.4), and our answer is shown in Sec. III.

III. CONSTRUCTING THE GENERATOR OF GAUGE TRANSFORMATIONS

We are going to compare two trajectories of the same physical system corresponding to the same initial data, but with two different choices of the arbitrary functions, namely, u^k and $u^k + \bar{u}^k$ where $\bar{u}_k$ is assumed to be a small deviation from the original functions u^k. According to the discussion of the preceding section the corresponding physical states of the system are to be considered as equivalent and so related by a gauge transformation with the generator of the form (2.4).

Let $Q = Q[q,p,Q_0,u]$ be any dynamical variable associated with the system and Q_0 its value for the initial data. (For simplicity we will assume that Q has no explicit time dependence, as this will not change the results.) We suppose that its evolution is generated by the total Hamiltonian so that, for the choice u^k of the arbitrary functions, we have

$$\dot{Q} = \{Q,H_T[u]\} = \{Q,H_c + u^k\phi_k\}. \tag{3.1}$$

Denoting $\bar{Q} = Q[q,p;Q_0,\ u+\bar{u}]$, we also have

$$\dot{\bar{Q}} = \{Q,H_T[u+\bar{u}]\} = \{Q,H_T[u] + \bar{u}^k\phi_k\}. \tag{3.2}$$

On the other hand, since Q and $\bar{Q}$ must be related by a gauge transformation it follows that

$$\bar{Q} = Q\,[u] + \{Q,G\}_{Q[u]} = Q\,[u] + \{Q,F^i(q,p)\phi_i\}_{Q[u]}. \tag{3.3}$$

The time evolution of $\bar{Q}$, as given by Eq. (3.3) above, is

$$\dot{\bar{Q}} = \{Q,H_T\} + F^i(q,p)\dot{\bar{\phi}}_i + F^i(q,p)\{\phi_i,H_T[u]\}.$$

Thus $\bar{Q}$ will be a solution of (3.2) if

$$\dot{F}^i(q,p)\bar{\phi}_i + \{Q,F^i(q,p)\{\phi_i,H_T[u]\},\} = \bar{u}^k\phi_k. \tag{3.4}$$

We now split the set of first class constraints into primary ϕ_i and secondary ψ_l ones, and write the generator G as

$$G = \omega^l(q,p)\psi_l + \varepsilon^n(q,p)\phi_n. \tag{3.5}$$

Equation (3.4) is then rewritten as

$$\frac{d\omega^l}{dt}\psi_l + \omega^l\{\psi_l,H_c\} + \omega^l u^n\{\psi_l,\phi_n\} + \varepsilon^n\{\phi_n,H_c\} = a_n\phi_n, \tag{3.6}$$

where in $a_n\phi_n$ we included all terms proportional to the primary first class constraints. Now, the quantities $\{\psi_l,H_c\}$ and $\{\phi_k,H_c\}$ in the above equation can be expressed as linear combinations of the secondary constraints. We write

$$\{\psi_l,H_c\} = \alpha_{ln}\psi_n, \quad \{\phi_n,H_c\} = \beta_{nl}\psi_l, \tag{3.7}$$

which, when substituted in (3.6), yields

$$\frac{d\omega^l}{dt}\psi_l + \omega^n\alpha_{nl}\psi_l + \varepsilon^n\beta_{nl}\psi_l + \omega^l u^n\{\psi_l,\phi_n\} = a_n\phi_n. \tag{3.8}$$

Taking into account the linear independence of the primary and secondary constraints, we see that if the last term on the left-hand side contains any linear combination of the secondary constraints, the coefficient ω^l will depend on the arbitrary functions u^k. If this is the case the generator G will lose its meaning, as it will not generate transformations between admissible trajectories. On the other hand, this term is equal to zero for the interesting physical systems known so far; thus we will discard it. Then we are left with the following differential equation relating the coefficients of the generator (3.5):

$$\frac{d\omega^l}{dt} + \alpha_{ln}\omega_n + \beta_{li}\varepsilon_i = 0. \tag{3.9}$$

This equation shows that the coefficients ω^l, ε^i are not independent of each other, so that given one set the other is determined by (3.9) provided that β_{li} and $\delta_{il}\, d/dt + \alpha_{il}$ are globally invertible. [In case one is looking for an explicit solution of (3.9) one must remember that the gauge transformations cannot change the initial data so that suitable initial conditions must be imposed.] We remark that we would not have obtained Eq. (3.9) if we had used the extended Hamiltonian as the generator of the dynamical evolution of the system.

IV. APPLICATIONS

(i) Let us consider the motion of a mechanical system in Euclidean space with coordinates $x(t)$, $y(t)$, $z(t)$, whose Lagrangian is given by

$$\mathscr{L} = \tfrac{1}{2}(\dot{x}^2 + \dot{y}^2) + \tfrac{1}{2}z^2(x^2 + y^2) - z(x\dot{y} - y\dot{x}) - \tfrac{1}{2}(x^2 + y^2), \tag{4.1}$$

where $\dot{x} = dx/dt$, etc. It is clear that this Lagrangian is singular since z is not a dynamical variable. By setting $z = 0$ one easily recognizes the remaining Lagrangian as that associated with a two-dimensional harmonic oscillator, invariant under time-independent rotations around the z axis. On the other hand, the symmetries of the Lagrangian (4.1) are not so obvious but can be revealed by the generator of gauge transformations.

With the canonical momenta defined by $p_x = \partial\mathscr{L}/\partial\dot{x}$, etc., one obtains

$$p_x = \dot{x} + xy, \quad p_y = \dot{y} - zx, \quad p_z = 0,$$

so that

$$\phi = p_z \approx 0 \tag{4.2}$$

is the primary constraint. Using the canonical Hamiltonian

$$H_c = \tfrac{1}{2}(p_x^2 + p_y^2) + z(xp_y - yp_x) + \tfrac{1}{2}(x^2 + y^2), \tag{4.3}$$

the consistency condition $\dot{\phi} = \{\phi, H_c\} \approx 0$ leads to the secondary constraint

$$\psi = yp_x - xp_y \approx 0. \tag{4.4}$$

The constraints (ϕ,ψ) are first class.

The generator of gauge transformation, Eq. (3.5), is simply

$$G = \omega(t)\psi + \varepsilon(t)\phi,$$

with

$$\dot{\omega} + \alpha\omega + \beta\varepsilon = 0. \tag{4.5}$$

The coefficients α and β are given by Eqs. (3.7), and one finds $\alpha = 0$, $\beta = 1$. Equation (4.5) reduces to $\dot{\omega} = -\varepsilon$ and the generator assumes the form

$$G = \omega\psi - \dot{\omega}\phi = \omega(yp_x - xp_y) - \dot{\omega}p_z. \tag{4.6}$$

It follows that

$$\delta x = \{x, G\} = \omega(t)y, \quad \delta y = -\omega(t)x, \quad \delta z = -\dot{\omega}(t). \tag{4.7}$$

The δx and δy as given above represent time-dependent (i.e., local) infinitesimal rotation by an angle $\omega(t)$ around the z axis. The transformation law for z is what should be expected for a gauge variable. It is easy to check that the Lagrangian (4.1) is invariant under (4.7).

(ii) Let us briefly consider the Yang–Mills field. We use the same notation and conventions as in Ref. 5. From the Lagrangian

$$\mathscr{L} = -\tfrac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a,$$

one obtains the primary constraints $\phi^a = \pi^a_0 \approx 0$. The consistency condition $\dot{\phi}^a = \{\phi^a, H_c\} \approx 0$, with

$$H_c = \int dx\left[\frac{1}{2}\pi^a_k\pi^a_k + \frac{1}{4}F^{ij}_a F^a_{ij} - A^0_a D^b_k \pi^k_b\right],$$

leads to the secondary constraints $\psi^a = D^k\pi^b_k \approx 0$. The generator now reads

$$G = \int d^3x(\omega^b\psi^b + \varepsilon^a\phi^a),$$

while Eqs. (3.9) and (3.7) are

$$\dot{\omega}^b + \varepsilon^a A^{ab} + \varepsilon^c B^{cb} = 0,$$

$$\{\psi^a, H_c\} = A^{ab}\psi^b, \quad \{\phi^a, H_0\} = B^{cb}\psi^b.$$

We find $A^{ab} = -gC^{acb}A^c_0$, $B^{cb} = \delta^{cb}$, and $\varepsilon^a = -D_0\omega^a$, where $D_0(\ )^a = \partial_0(\ )^a - gC^{abc}A^b_0(\ )^c$. The generator G assumes the form

$$G = \frac{1}{g}\int dz(\pi^\mu_a D_\mu\omega^a) \tag{4.8}$$

and it generates the Lagrangian gauge transformation law for the gauge potentials:

$$\delta A^b_\mu = \{G, A^b_\mu\} = (1/g)D_\mu\omega^b. \tag{4.9}$$

V. CONCLUSIONS

We made an analysis of some aspects of gauge transformations in the context of Dirac's theory of constrained systems. Accepting Dirac's conjecture that all first class constraints associated with a given physical system generate gauge transformations but using the total Hamiltonian to describe its dynamical evolution, we have been able to construct a generator for gauge transformations by comparing phase space trajectories with the same initial data but different choices of the arbitrary functions. The time evolution of the coefficients of the secondary constraints in our generator is not arbitrary but determined by the coefficients of the primary constraints. We observe that this does not mean any distinction between those constraints from the dynamical point of view. The generator we obtained obeys a closed algebra and leads to correct results when applied to the most familiar examples.

[1]P. A. M. Dirac, *Lectures in Quantum Mechanics* (Belfer Graduate School of Science, Monographs Series No. 2, Yeshiva Univ., New York, 1964).

[2]M. J. Gotay, J. Phys. A **16**, L141 (1983).
[3]R. L. Schafir, J. Phys. A **15**, L331 (1982).
[4]R. Sugano, Prog. Theor. Phys. **70**, 36 (1983).
[5]K. Sundermeyer, *Constrained Dynamics*, Lecture Notes in Physics, Vol. 169 (Springer-Verlag, New York, 1982).
[6]L. Castellani, Ann. Phys. (NY) **169**, 357 (1982).
[7]R. Di Stefano, Phys. Rev. D **27**, 1752 (1983).
[8]A. Hanson, T. Regge, and C. Teitelboim, *Constrained Hamiltonian Systems* (Accademia Nazionale dei Lincei, Rome, 1976).

Radial gauge in Poincaré gauge field theories

G. Modanese and M. Toller
Dipartimento di Fisica dell'Università, Trento, Italy and Istituto Nazionale di Fisica Nucleare, Gruppo Collegato di Trento, Trento, Italy

(Received 25 May 1989; accepted for publication 16 August 1989)

The generalization of the Fock–Schwinger or "radial" gauge condition $x^i A_i = 0$ to the gauge theories of the Poincaré group, which describe the gravitational field, is treated. It is shown that the choice of a radial gauge is equivalent to the use of normal coordinates and of tetrads parallel transported along autoparallel lines starting at the origin. The formulas that give the fields in the radial gauge starting from an arbitrary gauge and the formulas that give the gauge potentials in terms of the gauge field strengths are derived. The residual gauge freedom, which consists of the arbitrariness in the choice of the origin of the coordinate system and a tetrad of orthonormal vectors at this point is discussed in detail. It is the analog of the usual Poincaré invariance in flat space-time theories. The whole treatment can be extended to gauge theories of the affine and Euclidean groups. As an application, some properties of the homogeneous and isotropic states with random geometric fields are found.

I. INTRODUCTION

In classical gauge field theory it is often useful to impose gauge fixing conditions in such a way that the residual gauge freedom is reduced to the finite-dimensional group of the global gauge transformations, which do not depend on the space-time coordinates. A gauge fixing condition with this property is called "complete." Of course, one also requires that given an arbitrary configuration of the fields, there is at least one gauge transformation satisfying the gauge fixing conditions, which in this case are called "attainable."

For Yang–Mills or Maxwell fields, a simple condition that has both the properties described above is given by

$$x^i A_i(x) = 0, \tag{1.1}$$

where $A_i(x)$ are the gauge potentials and the indices of the gauge group are understood. This condition, called the Fock–Schwinger or coordinate gauge, has been known for a long time,[1,2] although it has rarely been used.[3–6] The condition is clearly invariant under homogeneous Lorentz transformations, but not under translations. The gauge potentials can be expressed univocally in terms of the field strengths F_{ik} by means of[3–8]

$$A_i(x) = \int_0^1 x^k F_{ki}(\lambda x)\lambda \, d\lambda. \tag{1.2}$$

An important feature of formula (1.2) is that it does not involve the behavior of the fields at infinity.

The aim of the present paper is to study a similar condition for a gauge theory of the Poincaré group, which describes the gravitational field.[9,10] In this case there are two kinds of gauge transformations: the general coordinate transformations (or the diffeomeorphisms of the space-time manifold) and the local Lorentz transformations of the tetrads $e_\alpha(x)$ which form an orthonormal basis in the tangent space of every point x. We also consider the dual basis given by the forms $e^\alpha(x)$ in the cotangent space of a point x.

A tensor can be described by its anholonomic components with respect to these tetrads (specified by Greek indices) or by its holonomic components (specified by Latin indices) with respect to the natural basis defined by the coordinate system x^i. We indicate by e^i_α and e^α_i the holonomic components of the tetrad vectors, namely we put

$$e_\alpha = e^i_\alpha \frac{\partial}{\partial x^i}, \quad e^\alpha = e^\alpha_i \, dx^i. \tag{1.3}$$

The components in (1.3) have the properties

$$e^i_\alpha e^\alpha_j = \delta^i_j, \quad e^i_\alpha e^\beta_i = \delta^\beta_\alpha \tag{1.4}$$

and can be used to transform the holonomic components of a tensor into anholonomic components and vice versa. For instance, for the metric tensor we have

$$g_{ij} = e^\alpha_i e^\beta_j g_{\alpha\beta} \tag{1.5}$$

and since the tetrads are assumed to be orthonormal, the anholonomic components $g_{\alpha\beta}$ are the same constants that appear in flat space theory.

The gauge potentials of a theory of this kind are the components $e^\alpha_i(x)$ of the dual tetrads and the anholonomic coefficients $\Gamma^\alpha_{\beta i}(x)$ of the connection, that appear in the covariant derivative D_i of a vector field given in terms of its anholonomic components V^α by

$$D_i V^\alpha = \partial_i V^\alpha + \Gamma^\alpha_{\beta i} V^\beta. \tag{1.6}$$

The dimension of the space-time and the signature of the metric are not relevant for our treatment, which also holds for gauge theories of the Euclidean group. Moreover, most of the following considerations do not involve the metric tensor and maintain their validity in a purely affine theory, namely in a gauge theory of the affine group,[11] where arbitrary tetrads of linearly independent vectors are admitted and the local transformations belong to the linear group. We use the term "homogeneous group" for the Lorentz, orthogonal, or linear group of the local transformations of the tetrads. We shall treat the general case and indicate the few points where the metric structure is relevant. For instance, for Poincaré or Euclidean gauge theories we require that the connection is of the metric type, namely that

$$-D_i g_{\alpha\beta} = \Gamma^\gamma_{\alpha i} g_{\gamma\beta} + \Gamma^\gamma_{\beta i} g_{\gamma\alpha} = 0, \tag{1.7}$$

 0022-2488/90/020452-07$03.00

but we do not use this condition in the main part of our treatment.

In Sec. II we formulate the gauge fixing conditions, similar to Eq. (1.1), which define the radial gauge: We discuss their geometric meaning and give the equations that determine the coordinates and tetrads in the radial gauge starting from an arbitrary gauge. The result is that the coordinates are normal[12] and the tetrads are parallel transported along autoparallel lines starting at the origin. The formalism obtained in this way can be considered as a special case of the path-dependent field theory developed by Mandelstam,[13–15] in which the paths are the autoparallel lines mentioned above.

The only arbitrariness in the construction of the radial gauge is the choice of the origin of the coordinate system and the tetrad at this point. This arbitrariness is interpreted as the residual gauge freedom. The treatment is local, namely the radial gauge is constructed in a suitable neighborhood of an arbitrary point. The corresponding global problem is much more difficult and cannot be solved in general. From this discussion it follows that the proposed gauge conditions are complete and attainable in a local sense and with the given definition of the residual gauge freedom.

In Sec. III we treat the analog of Eq. (1.2) and in Sec. IV we study the residual gauge transformations, which are considered as a generalization of the Poincaré transformations of Minkowski space-time. We give the formulas for finite displacements of the origin and the simpler formulas that describe infinitesimal displacements. In Sec. V we use these formulas for the description of homogeneous and isotropic states, namely states that are invariant under all the residual gauge transformations. This approach is useful for the treatment of states defined by random fields.

In the context of relativistic theories of gravitation, the results of the present paper can be applied not only to the pseudo-Riemannian four-dimensional space-time, but also to the three-dimensional spacelike surface which appears in the Hamiltonian treatment of general relativity.[16,17] The relation between quantum fields on Minkowski space and random classical fields on Euclidean space[18,19] has no well-established counterpart in the case of a curved space-time. However, we think that the study of random geometric and matter fields on a Riemannian manifold may provide very useful experience for the construction of a quantum theory of gravitation.

II. THE RADIAL GAUGE CONDITION

The radial gauge conditions we want to study are analogous to Eq. (1.1) and have the form

$$\xi^i \Gamma^\alpha_{\beta i}(\xi) = 0, \tag{2.1}$$

$$\xi^i e^\alpha_i(\xi) = \xi^i \delta^\alpha_i. \tag{2.2}$$

We have indicated by ξ^i the particular coordinate system that satisfies the gauge conditions in order to distinguish it from the arbitrary coordinate system x^i given initially. Condition (2.2) shows that it is consistent to use arbitrarily Latin or Greek indices to indicate the coordinates ξ.

We assume that all the fields are differentiable. By differentiation of Eqs. (2.1) and (2.2), we obtain

$$\Gamma^\alpha_{\beta i}(0) = 0, \quad e^\alpha_i(0) = \delta^\alpha_i. \tag{2.3}$$

Equations (2.3) show that in the gauge we are considering, the gauge potentials at the origin take the values they have in a flat space. In other words, it is possible to eliminate the gravitational field at a given point. This is a formulation of the equivalence principle, which is also valid in the presence of torsion when it is not possible to eliminate the holonomic connection coefficients at a given point.[20]

The holonomic connection coefficients, which appear in the covariant derivative of a vector written in terms of holonomic components, are given by

$$\Gamma^k_{ji} = e^k_\alpha e^\beta_j \Gamma^\alpha_{\beta i} + e^k_\alpha\, \partial_i\, e^\alpha_j \tag{2.4}$$

and condition (2.1) takes the more complicated form

$$\xi^i \Gamma^k_{ji} = e^k_\alpha \xi^i\, \partial_i\, e^\alpha_j. \tag{2.5}$$

From conditions (2.5) and (2.2) we obtain

$$\xi^i \xi^j\, \Gamma^k_{ji} = 0. \tag{2.6}$$

Now we show that given a point P_0 that fixes the origin and a tetrad $e_\alpha(P_0)$, conditions (2.1) and (2.2) determine univocally both the coordinate system ξ^i and the tetrad field in a neighborhood of P_0. This means that these gauge fixing conditions are locally attainable and complete. From the geometric point of view, condition (2.1) means that the tetrad $e_\alpha(\xi)$ is obtained from the given tetrad $e_\alpha(P_0)$ by parallel transport along the line

$$\xi^i(\lambda) = \lambda v^i, \quad \lambda \geqslant 0, \tag{2.7}$$

where the coefficients v^i are constant. In fact, if $u(\lambda) = u^\alpha e_\alpha(\xi(\lambda))$ is a vector with constant anholonomic components u^α we have

$$\frac{Du^\alpha}{d\lambda} = v^i \Gamma^\alpha_{\beta i}(\xi(\lambda)) u^\beta = 0. \tag{2.8}$$

The vector $v(\lambda)$ tangent to the curve (2.7) has constant holonomic components v^i proportional to $\xi^i(\lambda)$: From Eq. (2.2) we see that its anholonomic components with respect to the tetrads e_α are constant as well since they are given by

$$v^\alpha = v^i e^\alpha_i = v^i \delta^\alpha_i. \tag{2.9}$$

It follows that the tangent vector $v(\lambda)$ is parallel transported along the line (2.7). In conclusion, we have seen that the line (2.7) is autoparallel; the same result can be obtained by showing that as a consequence of condition (2.6), Eq. (2.7) satisfies the well-known differential equation

$$\frac{d^2\xi^k}{d\lambda^2} + \Gamma^k_{ji} \frac{d\xi^i}{d\lambda} \frac{d\xi^j}{d\lambda} = 0. \tag{2.10}$$

Remember that if the connection is metric and torsionless the autoparallel lines coincide with the geodesics.

As is well known, if the lines defined by Eq. (2.7) are autoparallel, the ξ^i are called normal coordinates.[12] Given the tetrad $e^\alpha(P_0)$, the ξ^i are uniquely determined in a neighborhood of P_0, where they form a chart of the space-time manifold. Also, the tetrads $e_\alpha(\xi)$ obtained by parallel transport are uniquely determined.

In order to give explicit formulas for the calculation of the normal coordinates and parallel-transported tetrads, we consider an arbitrary coordinate system x^i and indicate by x^i_0

the coordinates of the point P_0. We assume that the holonomic coefficients of the connection $\widehat{\Gamma}^i_{jk}(x)$ are known. We indicate by $\hat{e}^i_\alpha$ the components of the tetrads with respect to the natural holonomic basis determined by the coordinates x^i. The components of the dual tetrads in the natural basis determined by the normal coordinates ξ^i can be computed by means of

$$e^\alpha_i(\xi) = \frac{\partial x^k(\xi)}{\partial \xi^i}\, \hat{e}^\alpha_k(\xi). \tag{2.11}$$

Remember that the components of the tetrads and dual tetrads are connected by Eq. (1.4).

Since the tetrads are parallel transported along the line (2.7), we have

$$\frac{d\hat{e}^i_\beta(\lambda)}{d\lambda} = -\widehat{\Gamma}^i_{mn}(x(\lambda))\hat{e}^m_\beta(\lambda)\hat{e}^n_\alpha(\lambda)v^\alpha \tag{2.12}$$

and the fact that the vector with constant anholonomic components v^α is tangent to the line (2.7) is expressed by

$$\frac{dx^i(\lambda)}{d\lambda} = \hat{e}^i_\alpha v^\alpha. \tag{2.13}$$

Equations (2.12) and (2.13), with the initial conditions $x^i(0) = x^i_0$ and $\hat{e}^i_\alpha(0) = e^i_\alpha(P_0)$, determine the quantities x^i and $\hat{e}^i_\beta$ as functions of λ and v^α: However, it is easy to see that x^i and $\hat{e}^i_\beta$ depend on a particular combination of these variables, namely on the normal coordinates $\xi^\alpha = \lambda v^\alpha$. If we are considering a metric space and a metric connection, we can choose an orthonormal initial tetrad $e_\alpha(0)$ and all the parallel-transported tetrads are automatically orthonormal.

It is convenient to transform the differential equations (2.12) and (2.13) with their initial conditions into the following pair of coupled integral equations:

$$x^i(\xi) = x^i_0 + \xi^\alpha \int_0^1 \hat{e}^i_\alpha(\lambda\xi)d\lambda, \tag{2.14}$$

$$\hat{e}^i_\beta(\xi) = \hat{e}^i_\beta(0) - \xi^\alpha \int_0^1 \widehat{\Gamma}^i_{mn}(x(\lambda\xi))\hat{e}^m_\beta(\lambda\xi)\hat{e}^n_\alpha(\lambda\xi)d\lambda. \tag{2.15}$$

It is possible to solve Eqs. (2.14) and (2.15) perturbatively to any desired order in $\widehat{\Gamma}$ by substituting at each step the lower order solution in the lhs integrals.

Note that if we change the initial conditions by performing a Lorentz transformation L of the tetrad $e_\alpha(0)$, we find a new solution obtained from the old one by means of the same Lorentz transformation L applied to all the tetrads and normal coordinates. This is part of the residual global gauge transformations; the other part is determined by the arbitrary choice of the origin P_0 and affects the normal coordinates and parallel-transported tetrads in a more complicated way, which we shall discuss in Sec. IV.

The radial gauge conditions (2.1) and (2.2) can be regarded in a sense as an operational prescription which permits one to locate the measuring instruments in a neighborhood of the observer, who lies at the origin P_0. In fact, a simple way to explore this neighborhood is to send from the origin many "space probes" carrying clocks, gyroscopes, and measuring instruments. A space probe will be launched with four-velocity v^α with respect to the given tetrad $e_\alpha(0)$ and, if τ is the proper time measured by the clock, $\xi^\alpha = \tau v^\alpha$ are the normal coordinates (in the absence of torsion). Of course, only the interior of the future cone can be explored in this way. This operational interpretation presents some difficulties which need a deeper analysis if gravitation and quantum theory are taken into account at the same time. In fact, the mass of the probe has to be small in order to avoid perturbations of the gravitational field and has to be large in order to minimize the effects of the uncertainty relations. Moreover, a probe with a nonvanishing mass must have an extension larger than the Schwarzschild radius and is influenced by the average value of the gravitational field in a finite region. A space-time average of this kind is also necessary in order to avoid a divergence of the quantum fluctuations of the gravitational field.

III. CALCULATION OF THE GAUGE POTENTIALS FROM THE CURVATURES

In this section we derive from the radial gauge conditions (2.1) and (2.2) two formulas analogous to Eq. (1.2) which give the field potentials $\Gamma^\alpha_{\beta i}(\xi)$ and $e^\alpha_i(\xi)$ in terms of the Riemann tensor $R^{\ \alpha}_{\beta ik}(\xi)$ and the torsion tensor $S^{\ \alpha}_{ik}(\xi)$. We remember that these tensors are given by

$$R^{\ \alpha}_{\beta ik} = \partial_i\,\Gamma^\alpha_{\beta k} - \partial_k\,\Gamma^\alpha_{\beta i} + \Gamma^\alpha_{\gamma i}\Gamma^\gamma_{\beta k} - \Gamma^\alpha_{\gamma k}\Gamma^\gamma_{\beta i}, \tag{3.1}$$

$$S^{\ \alpha}_{ik} = \partial_i\, e^\alpha_k - \partial_k\, e^\alpha_i + e^\beta_k\Gamma^\alpha_{\beta i} - e^\beta_i\Gamma^\alpha_{\beta k}. \tag{3.2}$$

Let us multiply Eq. (3.1) by ξ^i; from condition (2.1) we have

$$\xi^i R^{\ \alpha}_{\beta ik}(\xi) = \xi^i\partial_i\,\Gamma^\alpha_{\beta k}(\xi) + \Gamma^\alpha_{\beta k}(\xi). \tag{3.3}$$

If we now put $\xi \to \lambda\xi$ we obtain

$$\frac{d}{d\lambda}(\lambda\Gamma^\alpha_{\beta k}(\lambda\xi)) = \lambda\xi^i R^{\ \alpha}_{\beta ik}(\lambda\xi); \tag{3.4}$$

integrating, we finally have

$$\Gamma^\alpha_{\beta k}(\xi) = \xi^i \int_0^1 R^{\ \alpha}_{\beta ik}(\lambda\xi)\lambda\, d\lambda. \tag{3.5}$$

In a similar way, by multiplying Eq. (3.2) by ξ^i and taking into account condition (2.2) we obtain

$$\xi^i S^{\ \alpha}_{ik}(\xi) = \xi^i\partial_i e^\alpha_k(\xi) + e^\alpha_k(\xi) - \delta^\alpha_k - \xi^\beta\Gamma^\alpha_{\beta k}(\xi); \tag{3.6}$$

by the same procedure we obtain

$$\frac{d}{d\lambda}(\lambda(e^\alpha_k(\lambda\xi) - \delta^\alpha_k)) = \lambda\xi^\beta\Gamma^\alpha_{\beta k}(\lambda\xi) + \lambda\xi^i S^{\ \alpha}_{ik}(\lambda\xi) \tag{3.7}$$

and

$$e^\alpha_k(\xi) = \delta^\alpha_k + \int_0^1 [\xi^\beta\Gamma^\alpha_{\beta k}(\lambda\xi) + \xi^i S^{\ \alpha}_{ik}(\lambda\xi)]\lambda\, d\lambda. \tag{3.8}$$

By substituting Eq. (3.5) into Eq. (3.8) we have

$$e^\alpha_k(\xi) = \delta^\alpha_k + \xi^i\xi^\beta \int_0^1 R^{\ \alpha}_{\beta ik}(\lambda\xi)(1-\lambda)\lambda\, d\lambda + \xi^i \int_0^1 S^{\ \alpha}_{ik}(\lambda\xi)\lambda\, d\lambda. \tag{3.9}$$

Equations (3.5) and (3.9) are the analogs of Eq. (1.2): The gauge potentials they give satisfy the gauge conditions

(2.1) and (2.2) as a consequence of the antisymmetry of $R^{\ \alpha}_{\beta ik}$ and $S^{\ \alpha}_{ik}$ with respect to the indices i, k. However, Eqs. (3.5) and (3.9) solve (3.1) and (3.2) only if the functions $R^{\ \alpha}_{\beta ik}(\xi)$ and $S^{\ \alpha}_{ik}(\xi)$ satisfy some conditions. These conditions have been derived and used in the Yang–Mills case[5,7] and a similar treatment can be given in the case we are considering. If we substitute Eqs. (3.5) and (3.9) into Eqs. (3.1) and (3.2), after a long calculation we find that they are equivalent to the following projected Bianchi identities:

$$\xi^i \sum_{\{ijk\}} (\partial_i R^{\ \alpha}_{\beta jk} + \Gamma^\alpha_{\gamma i} R^{\ \gamma}_{\beta jk} - \Gamma^\gamma_{\beta i} R^{\ \alpha}_{\gamma jk}) = 0, \qquad (3.10)$$

$$\xi^i \sum_{\{ijk\}} (\partial_i S^{\ \alpha}_{jk} + \Gamma^\alpha_{\beta i} S^{\ \beta}_{jk} - e^\beta_i R^{\ \alpha}_{\beta jk}) = 0, \qquad (3.11)$$

in which the potentials e^β_i and $\Gamma^\alpha_{\beta i}$ are replaced by the integral expressions (3.5) and (3.9). We have indicated by $\Sigma_{\{ijk\}}$ the sum over the cyclic permutations of the indices i, j, k. These conditions have a nonlocal character since they are expressed by integrodifferential equations.

The metricity condition (1.7) is satisfied by expression (3.5) if the curvature satisfies the analogous local condition

$$g_{\alpha\gamma} R^{\ \gamma}_{\beta ik} = -g_{\beta\gamma} R^{\ \gamma}_{\alpha ik}. \qquad (3.12)$$

The Einstein field equations $e^i_\alpha R^{\ \alpha}_{\beta ik} = 0$ contain the potential e^i_α and therefore take a nonlocal character if we want to express them in terms of the curvature alone. In conclusion, if we try to use Eqs. (3.5) and (3.9) to eliminate the gauge potentials from the theory, we get very complicated nonlocal field equations.

The situation is not much better if we use the holonomic components $R^{\ m}_{nik}(\xi)$ and $S^{\ m}_{ik}(\xi)$ or the completely anholonomic components $R^{\ \alpha}_{\beta\gamma\delta}(\xi)$ and $S^{\ \alpha}_{\gamma\delta}(\xi)$. We have to modify Eqs. (3.5) and (3.9), introducing in the rhs some tetrads with suitable indices. As a result we obtain formulas which, unlike Eqs. (3.5) and (3.9), are not simple integral expressions, but integral equations of the Volterra kind.

If the components of curvature and torsion can be expanded in power series of the normal coordinates, Eqs. (3.5) and (3.9) can be integrated term by term and give power expansions of the potentials with coefficients expressed in terms of curvature, torsion, and their covariant derivatives at the origin $\xi = 0$.

IV. RESIDUAL GAUGE TRANSFORMATIONS

If we adopt the radial gauge, the geometry of the space-time manifold in a neighborhood of P_0 is completely described by the functions $e^k_\alpha(\xi)$, $\Gamma^\alpha_{\beta k}(\xi)$, which have to satisfy the gauge conditions (2.1) and (2.2), possibly the metricity condition (1.7), and the field equations of the theory we are considering. In a similar way, the matter fields are completely described by their components with respect to the tetrads, expressed as functions of the normal coordinates ξ. The only arbitrary elements in this description are the choice of the origin P_0 and tetrad $e_\alpha(P_0)$. This is the residual gauge freedom, which remains when we impose the radial gauge conditions.

Since only gauge-invariant quantities are observable and the changes of the origin P_0 and tetrad $e_\alpha(P_0)$ are considered as gauge transformations, we may ask if the fields in the radial gauge can be considered as observable quantities: The answer is that they are as observable as the fields in Minkowski space-time described by their components with respect to a given inertial coordinate frame. In fact, in this case as well the values of the field components depend on the choice of the origin and the directions of the coordinate axes. The transformation properties of the fields with respect to the Poincaré group permit us to compute the components in the new reference frame in terms of the old components. For instance, if V^α is a vector field, we have

$$V'^\alpha(x') = [L^{-1}]^\alpha_\beta V^\beta(x), \quad x = Lx' + x_0. \qquad (4.1)$$

In the following we generalize formula (4.1) to a curved space-time with radial gauge, namely we derive the explicit form of the residual gauge transformations. We consider a radial gauge with origin at the point P_0, coordinates ξ^i, and tetrads $e_\alpha(\xi)$ and we start from the point P_1 with coordinates ξ^i_1 and the tetrad

$$e'_\alpha(0) = L^{\ \beta}_\alpha e_\beta(\xi_1) \qquad (4.2)$$

to build a new radial gauge with coordinates ξ'^i and tetrads $e'_\alpha(\xi')$. If, for simplicity, we take $L = 1$, the coordinates and tetrads are connected by

$$\xi^i = \Xi^i(\xi_1,\xi'), \quad e'_\alpha(\xi') = \Omega^\beta_\alpha(\xi_1,\xi')e_\beta(\xi). \qquad (4.3)$$

It is easy to see that for general values of L relations (4.3) are modified as follows:

$$\xi^i = \Xi^i(\xi_1,L\xi'), \quad e'_\alpha(\xi') = L^{\ \beta}_\alpha \Omega^\gamma_\beta(\xi_1,L\xi')e_\gamma(\xi). \qquad (4.4)$$

The transformation property of a vector field is

$$V'^\alpha(\xi') = [\Omega^{-1}(\xi_1,L\xi')L^{-1}]^\alpha_\beta V^\beta(\xi); \qquad (4.5)$$

tensors of arbitrary order transform in a similar way. In Poincaré or Euclidean gauge theories it is also easy to write the transformation properties of spinor fields. In a flat space-time we have

$$\Xi^i(\xi_1,\xi') = \xi^i_1 + \xi'^i, \quad \Omega^\beta_\alpha(\xi_1,\xi') = \delta^\beta_\alpha \qquad (4.6)$$

and the transformation property (4.5) takes the form (4.1). It follows from the definitions that Eq. (4.6) holds for general spaces when ξ_1 and ξ' are proportional.

The gauge potentials, which describe the geometry, transform in the following way (for $L = 1$):

$$e'^\alpha_i(\xi') = \frac{\partial \Xi^k(\xi_1,\xi')}{\partial \xi'^i}[\Omega^{-1}(\xi_1,\xi')]^\alpha_\beta e^\beta_k(\xi), \qquad (4.7)$$

$$\Gamma'^\alpha_{\beta i}(\xi') = [\Omega^{-1}(\xi_1,\xi')]^\alpha_\gamma \left[\Omega^\delta_\beta(\xi_1,\xi')\frac{\partial \Xi^k(\xi_1,\xi')}{\partial \xi'^i} \times \Gamma^\gamma_{\delta k}(\xi) + \frac{\partial \Omega^\gamma_\beta(\xi_1,\xi')}{\partial \xi'^i}\right]. \qquad (4.8)$$

Note that the residual gauge transformations do not form a group because the quantities $\Xi^i(\xi_1,\xi')$ and $\Omega^\beta_\alpha(\xi_1,\xi')$ and also depend on the initial point P_0: They can be computed by means of the method given in Sec. II. We remark that the quantities Ω^β_α are the components of the tetrads e'_α in the anholonomic basis defined by e_β and that the vectors e'_α are parallel transported along the line

$$\xi'^i = \lambda v^i. \qquad (4.9)$$

Therefore, we have

$$\frac{D\Omega^{\beta}_{\alpha}}{d\lambda}=\frac{d\Omega^{\beta}_{\alpha}}{d\lambda}+\frac{d\Xi^{k}}{d\lambda}\Gamma^{\beta}_{\gamma k}\Omega^{\gamma}_{\alpha}=0. \tag{4.10}$$

The vectors tangent to the line (4.9) have constant components v^α with respect to the tetrads e'_α, namely

$$\frac{d\Xi^{k}}{d\lambda}=v^{\alpha}\Omega^{\beta}_{\alpha}e^{k}_{\beta}. \tag{4.11}$$

From Eqs. (4.10) and (4.11), we obtain

$$\frac{d\Omega^{\beta}_{\alpha}}{d\lambda}=-v^{\delta}\Omega^{\epsilon}_{\delta}e^{k}_{\epsilon}\Gamma^{\beta}_{\gamma k}\Omega^{\gamma}_{\alpha}. \tag{4.12}$$

The differential equations (4.11) and (4.12) with suitable initial conditions determine the quantities Ξ and Ω: It is convenient to transform them into the following integral equations:

$$\Xi^{k}(\xi_1,\xi')=\xi^{k}_{1}+\xi'^{\alpha}\int_0^1\Omega^{\beta}_{\alpha}(\xi_1,\lambda\xi')e^{k}_{\beta}(\Xi(\xi_1,\lambda\xi'))d\lambda, \tag{4.13}$$

$$\Omega^{\beta}_{\alpha}(\xi_1,\xi')=\delta^{\beta}_{\alpha}-\xi'^{\delta}\int_0^1\Omega^{\epsilon}_{\delta}(\xi_1,\lambda\xi')e^{k}_{\epsilon}(\Xi(\xi_1,\lambda\xi'))\times\Gamma^{\beta}_{\gamma k}(\Xi(\xi_1,\lambda\xi'))\Omega^{\gamma}_{\alpha}(\xi_1,\lambda\xi')d\lambda. \tag{4.14}$$

It is useful to consider an infinitesimal displacement ξ_1 of the origin and put

$$\Xi^{k}(\xi_1,\xi')=\xi^{k}_{1}+\xi'^{k}+\xi^{i}_{1}A^{k}_{i}(\xi')+O(\xi^{2}_{1}), \tag{4.15}$$

$$\Omega^{\beta}_{\alpha}(\xi_1,\xi')=\delta^{\beta}_{\alpha}+\xi^{i}_{1}B^{\beta}_{\alpha i}(\xi')+O(\xi^{2}_{1}). \tag{4.16}$$

From Eq. (4.6), which holds when ξ_1 and ξ' are proportional, we obtain the conditions

$$\xi^{i}A^{k}_{i}(\xi)=0,\quad \xi^{i}B^{\beta}_{\alpha i}(\xi)=0. \tag{4.17}$$

Formulas (4.5), (4.7), and (4.8) take the form

$$\begin{aligned}
&V'^{\alpha}(\xi)-V^{\alpha}(\xi)=\xi^{i}_{1}\delta_{j}V^{\alpha}+O(\xi^{2}_{1}),\\
&e'^{\alpha}_{i}(\xi)-e^{\alpha}_{i}(\xi)=\xi^{j}_{1}\delta_{j}e^{\alpha}_{i}+O(\xi^{2}_{1}),\\
&\Gamma'^{\alpha}_{\beta i}(\xi)-\Gamma^{\alpha}_{\beta i}(\xi)=\xi^{j}_{1}\delta_{j}\Gamma^{\alpha}_{\beta i}+O(\xi^{2}_{1}),
\end{aligned} \tag{4.18}$$

where we have introduced the definitions

$$\delta_{j}V^{\alpha}=-B^{\alpha}_{\beta j}V^{\beta}+(\delta^{k}_{j}+A^{k}_{j})\frac{\partial V^{\alpha}}{\partial\xi^{k}}, \tag{4.19}$$

$$\delta_{j}e^{\alpha}_{i}=-B^{\alpha}_{\beta j}e^{\beta}_{i}+\frac{\partial A^{k}_{j}}{\partial\xi^{i}}e^{\alpha}_{k}+(\delta^{k}_{j}+A^{k}_{j})\frac{\partial e^{\alpha}_{i}}{\partial\xi^{k}}, \tag{4.20}$$

$$\delta_{j}\Gamma^{\alpha}_{\beta i}=-B^{\alpha}_{\gamma j}\Gamma^{\gamma}_{\beta i}+B^{\gamma}_{\beta j}\Gamma^{\alpha}_{\gamma i}+\frac{\partial A^{k}_{j}}{\partial\xi^{i}}\Gamma^{\alpha}_{\beta k}+(\delta^{k}_{j}+A^{k}_{j})\frac{\partial\Gamma^{\alpha}_{\beta i}}{\partial\xi^{k}}+\frac{\partial B^{\alpha}_{\beta j}}{\partial\xi^{i}}. \tag{4.21}$$

Note that in expressions (4.19)–(4.21) one can recognize terms related to the anholonomic indices, terms related to the holonomic indices, and terms with the partial derivatives of the field. One can easily imagine analogous formulas for other tensor fields. Equation (4.21) has one more term, which results from the nontensor character of the connection coefficients.

If we substitue expressions (4.15) and (4.16) into Eqs. (4.13) and (4.14) and disregard second-order terms, we obtain the following integral equations for the quantities A and B:

$$A^{k}_{i}(\xi)=\int_0^1(A^{j}_{i}(\lambda\xi)+\delta^{j}_{i})\delta^{\gamma}_{j}(\delta^{k}_{\gamma}-e^{k}_{\gamma}(\lambda\xi))\lambda^{-1}d\lambda+\xi^{\gamma}\int_0^1B^{\delta}_{\gamma i}(\lambda\xi)e^{k}_{\delta}(\lambda\xi)d\lambda, \tag{4.22}$$

$$B^{\beta}_{\alpha i}(\xi)=\int_0^1(A^{j}_{i}(\lambda\xi)+\delta^{j}_{i})\delta^{\gamma}_{j}e^{k}_{\gamma}(\lambda\xi)\Gamma^{\beta}_{\alpha k}(\lambda\xi)\lambda^{-1}d\lambda-\xi^{\gamma}\int_0^1B^{\delta}_{\gamma i}(\lambda\xi)e^{k}_{\delta}(\lambda\xi)\Gamma^{\beta}_{\alpha k}(\lambda\xi)d\lambda. \tag{4.23}$$

We have used the following consequences of the gauge conditions (2.1) and (2.2):

$$\begin{aligned}
&\xi^{\alpha}\partial_{j}e^{k}_{\alpha}(\xi)=\delta^{k}_{j}-\delta^{\alpha}_{j}e^{k}_{\alpha}(\xi),\\
&\xi^{\delta}\partial_{j}(e^{k}_{\delta}(\xi)\Gamma^{\beta}_{\alpha k}(\xi))=-\delta^{\gamma}_{j}e^{k}_{\gamma}(\xi)\Gamma^{\beta}_{\alpha k}(\xi).
\end{aligned} \tag{4.24}$$

For small values of $e^{i}_{\alpha}-\delta^{i}_{\alpha}$ and $\Gamma^{\beta}_{\alpha i}$, we obtain the first-order perturbative solution

$$\begin{aligned}
A^{k}_{i}(\xi)&\approx\delta^{\gamma}_{i}\int_0^1(\delta^{k}_{\gamma}-e^{k}_{\gamma}(\lambda\xi))\lambda^{-1}d\lambda+\delta^{k}_{\delta}\xi^{\gamma}\int_0^1\Gamma^{\delta}_{\gamma i}(\lambda\xi)(\lambda^{-1}-1)d\lambda\\
&=\delta^{k}_{\gamma}\xi^{j}\int_0^1S^{\gamma}_{ji}(\lambda\xi)(1-\lambda)d\lambda+\delta^{k}_{\gamma}\xi^{\beta}\xi^{j}\int_0^1R^{\gamma}_{\beta ji}(\lambda\xi)(1-\lambda)^2d\lambda,
\end{aligned} \tag{4.25}$$

$$B^{\beta}_{\alpha i}(\xi)\approx\int_0^1\Gamma^{\beta}_{\alpha i}(\lambda\xi)\lambda^{-1}d\lambda=\xi^{k}\int_0^1R^{\beta}_{\alpha ki}(\lambda\xi)(1-\lambda)d\lambda. \tag{4.26}$$

The last equalities in Eqs. (4.25) and (4.26) have been obtained by means of Eqs. (3.5) and (3.9).

V. HOMOGENEOUS AND ISOTROPIC RANDOM FIELDS

We have seen in Sec. IV how the physical observables are affected by the residual gauge transformations, which play a role analogous to the Poincaré transformations in flat space-time theories. It is interesting to consider states in which all the observables are invariant under residual gauge transformations: They are the analog of the vacuum state of flat space-time theories and are isotropic and homogeneous (in space-time) since their description does not depend on the position, time, orientation, and velocity of the observer. Similar considerations are relevant for the gauge theories of the affine or Euclidean groups in arbitrary dimension.

In classical field theory, as we shall see, the homogeneous isotropic "pure" state are given by flat or constant curvature manifolds whose properties are known from elementary geometry. On the other hand, the results of the preceding sections cannot be applied directly to a quantum field theory for at least two reasons: First, we have no argument to determine the order of the noncommuting factors; moreover, quantum fields are distributions and require a more careful treatment.

An interesting application of our results concerns classical "mixed" states, namely classical random fields; they can describe a random superposition of gravitational waves. However, we shall see that in a Poincaré gauge theory, the only homogeneous and isotropic nonsingular states of this kind are nonfluctuating "pure" states. A random field could also describe the vacuum fluctuations in a hypothetical Euclidean counterpart of a quantum theory of gravitation or the quantum fluctuations of a set of quantum fields which commute on a spacelike surface. However, the last two applications should take into account the singular nature of the fields; as a consequence, the following considerations can be considered only as a preliminary exercise with respect to these physically relevant problems.

We indicate by $\phi_1(\xi_1),...,\phi_n(\xi_n)$ a set of fields, possibly including the gauge potentials e_i^α and $\Gamma^\alpha_{\beta i}$. The indices, holonomic or anholonomic, are understood. A random state can be described by means of the averages

$$\langle \phi_1(\xi_1)\cdots\phi_n(\xi_n)\rangle. \tag{5.1}$$

We assume that these expressions are differentiable functions of the variables $\xi_1,...,\xi_n$; this is our definition of nonsingular random fields. The invariance of these quantities with respect to the homogeneous (Lorentz, linear, or orthogonal) transformations of the tetrad $e_\alpha(P_0)$ is imposed as in a flat theory; the quantities do not change when one operates on all the holonomic and anholonomic field indices and the coordinates ξ^i with a matrix L of the homogeneous group.

The simplest consequence is that the average $\langle\phi(0)\rangle$ is an invariant tensor which has the same components in all the allowed reference frames. This holds in particular for the curvature and torsion tensors: Their average at the origin must vanish in an affine theory; in a gauge theory of the Poincaré or the Euclidean groups we must have

$$\langle S^{\alpha}_{ik}(0)\rangle = 0, \quad \langle R^{\alpha\beta}_{ik}(0)\rangle = r(\delta^\alpha_i\delta^\beta_k - \delta^\alpha_k\delta^\beta_i). \tag{5.2}$$

In the three-dimensional case we also have to consider the possibility

$$\langle S^{\alpha}_{ik}(0)\rangle = r'\epsilon^\alpha{}_{ik}, \tag{5.3}$$

where $\epsilon^{\alpha\beta\gamma}$ is the unit antisymmetric tensor.

The invariance with respect to infinitesimal translations, namely the infinitesimal parallel displacements of the tetrad $e_\alpha(P_0)$, has the more complicated form

$$\delta_j\langle\phi_1(\xi_1)\cdots\phi_n(\xi_n)\rangle = \langle\delta_j\phi_1(\xi_1)\cdots\phi_n(\xi_n)\rangle + \cdots + \langle\phi_1(\xi_1)\cdots\delta_j\phi_n(\xi_n)\rangle = 0, \tag{5.4}$$

where the differential operator δ_j is defined by Eqs. (4.19)–(4.21). Condition (5.4) should be considered as important as the translational invariance in flat space-time theories. In the following we clarify its meaning by means of some applications.

The simplest consequence of Eq. (5.4) is

$$\xi^j\langle\delta_j V^\alpha(\xi)\rangle = 0. \tag{5.5}$$

From Eqs. (4.19) and (4.17) we obtain

$$\xi^j\partial_j\langle V^\alpha(\xi)\rangle = 0, \tag{5.6}$$

which gives

$$\langle V^\alpha(\xi)\rangle = \langle V^\alpha(0)\rangle. \tag{5.7}$$

The same conclusion holds for all the tensor fields which have only anholonomic indices: Their average is a constant tensor invariant under the homogeneous group. In particular, from Eqs. (5.2) we have

$$\langle S^{\alpha}_{\beta\gamma}(\xi)\rangle = 0, \quad \langle R^{\alpha\beta}_{\gamma\delta}(\xi)\rangle = r(\delta^\alpha_\gamma\delta^\beta_\delta - \delta^\alpha_\delta\delta^\beta_\gamma). \tag{5.8}$$

For the gauge potentials from Eqs. (4.20), (4.21), and (4.17) we obtain, in a similar way, the conditions

$$\xi^j\partial_j\langle e^\alpha_i\rangle = \langle A^k_i e^\alpha_k\rangle, \tag{5.9}$$

$$\xi^j\partial_j\langle\Gamma^\alpha_{\beta i}\rangle = \langle A^k_i\Gamma^\alpha_{\beta k}\rangle + \langle B^\alpha_{\beta i}\rangle, \tag{5.10}$$

which show that the averages of fields with holonomic indices have more complicated properties.

If we consider two vector fields at two points that lie on the same ray, from the condition

$$\xi^j\langle\delta_j V^\alpha(\xi)U^\beta(\lambda\xi)\rangle + \xi^j\langle V^\alpha(\xi)\delta_j U^\beta(\lambda\xi)\rangle = 0 \tag{5.11}$$

we obtain

$$\langle V^\alpha(\xi)U^\beta(\lambda\xi)\rangle = \langle V^\alpha((1-\lambda)\xi)U^\beta(0)\rangle. \tag{5.12}$$

In general, one can see that the average of a product of tensor fields with anholonomic indices on the same ray is invariant under translations along the ray.

In the absence of fluctuations from Eqs. (5.8) we obtain

$$S^{\alpha}_{ik}(\xi) = 0, \quad R^{\alpha\beta}_{ik}(\xi) = r(e^\alpha_i(\xi)e^\beta_k(\xi) - e^\alpha_k(\xi)e^\beta_i(\xi)). \tag{5.13}$$

If we substitute expressions (5.13) into Eqs. (3.4) and (3.7) we obtain a system of differential equations which can easily be solved and give the following explicit expressions for the gauge potentials of a space-time with constant curvature:

$$\Gamma^{\alpha\beta}_k(\xi) = s^{-1}(\cos\sqrt{rs} - 1)(\delta^\alpha_k\xi^\beta - \delta^\beta_k\xi^\alpha), \tag{5.14}$$

$$e^\alpha_k(\xi) = \delta^\alpha_k + (\sin\sqrt{rs}/\sqrt{rs} - 1)(\delta^\alpha_k - s^{-1}\xi^\alpha\xi_\beta\delta^\beta_k), \tag{5.15}$$

where $s = \xi^\alpha\xi_\alpha$. Note that the singularities at $rs = 0$ compensate and a simple analytic continuation is necessary for $rs < 0$.

Now we want to show that in a classical Poincaré gauge theory with nonsingular random fields, vacuum fluctuations are not permitted and the nonfluctuating constant curvature spaces described above are the only possible vacuum states. More precisely, we shall show that in a homogeneous isotropic state all the irreducible tensor fields vanish and the scalar fields are constant. A similar result does not hold in the Euclidean gauge theories, to which the results given above can be applied in a nontrivial way.

As a first example, we consider a vector field $V^\alpha(\xi)$ and the tensor

$$T^{\alpha\beta}(\xi) = V^\alpha(\xi)V^\beta(\xi). \tag{5.16}$$

Since the only invariant tensor of second order is the metric tensor, from the results obtained above we have

$$\langle T^{\alpha\beta}(\xi)\rangle = cg^{\alpha\beta}, \tag{5.17}$$

namely

$$\langle(V^0(\xi))^2\rangle = c, \quad \langle(V^\alpha(\xi))^2\rangle = -c, \quad \alpha = 1,2,3. \tag{5.18}$$

It follows that $c = 0$ and all the components of the vector field (not just their averages) vanish. Note that the negative values of $g^{\alpha\beta}$ are essential in order to obtain this result. Note, also, that we have assumed neither conservation laws nor field equations. The only assumption on the field is its transformation property and the existence of the averages (5.18). The last assumption is essential since it is easy to find Poincaré invariant singular random fields in Minkowski space-time.

By considering its gradient, we see that a scalar field must be a possibly fluctuating constant. If we assume that correlations decrease at large distances, this constant field cannot fluctuate. In order to complete our argument, we decompose an arbitrary tensor field into irreducible, possibly complex, fields $\Phi^A(\xi)$, where A indicates a set of anholonomic indices. If we exclude the scalar fields, which have already been considered, one of these fields transforms according to an irreducible faithful finite-dimensional representation $D^A_B(L)$ of the Lorentz group. According to our previous discussion, the matrix

$$G^{AB} = \langle \overline{\Phi}^A(\xi)\Phi^B(\xi)\rangle \tag{5.19}$$

is constant, positive semidefinite, and invariant with respect to the representation D. If G^{AB} does not vanish, it defines a scalar product, which makes D a unitary representation. Since the Lorentz group has no faithful finite-dimensional unitary representations, expression (5.19) vanishes and we have $\langle|\Phi^A(\xi)|^2\rangle = 0$.

[1]V. A. Fock, Sov. Phys. **12**, 404 (1937).
[2]J. Schwinger, Phys. Rev. **82**, 684 (1952).
[3]M. A. Shifman, Nucl. Phys. B **173**, 13 (1980).
[4]W. E. Brittin, W. Rodman Smythe, and W. Wyss, Am. J. Phys. **50**, 693 (1982).
[5]L. Durand and E. Mendel, Phys. Rev. D **26**, 1368 (1982).
[6]W. Kummer and J. Weiser, Z. Phys. C **31**, 105 (1986).
[7]C. Cronström, Phys. Lett. **90B**, 267 (1980).
[8]B. K. Skagerstam, Am. J. Phys. **51**, 1148 (1983).
[9]F. W. Hehl, P. von der Heyde, G. D. Kerlick, and J. M. Nester, Rev. Mod. Phys. **48**, 393 (1976).
[10]F. W. Hehl, in *Cosmology and Gravitation*, edited by P. G. Bergmann and V. De Sabbata (Plenum, New York, 1980).
[11]P. Baekler, F. W. Hehl, and E. W. Mielke, in *Proceedings of the Fourth Marcel Grossmann Meeting on General Relativity*, edited by R. Ruffini (North-Holland, Amsterdam, 1986), p. 277.
[12]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley, New York, 1969).
[13]S. Mandelstam, Ann. Phys. (NY) **19**, 25 (1962).
[14]S. Mandelstam, Phys. Rev. **175**, 1580 (1968).
[15]S. Mandelstam, Phys. Rev. **175**, 1604 (1968).
[16]A. Hanson, T. Regge, and C. Teitelboim, in *Constrained Hamiltonian Systems*, Lecture Notes (Accademia Nazionale dei Lincei, Rome, 1976).
[17]J. E. Nelson and T. Regge; *Proceedings of the VIII Italian Conference on General Relativity and Gravitational Physics*, edited by M. Cerdonio (World Scientific, Singapore, 1989).
[18]E. Nelson, J. Funct. Anal. **12**, 97 (1973).
[19]K. Osterwalder and R. Schrader, Commun. Math. Phys. **42**, 281 (1975).
[20]P. von der Heyde, Nuovo Cimento Lett. **14**, 250 (1975).

A supersymmetric Benney hierarchy: A semiintegrable system

B. A. Kupershmidt
The University of Tennessee Space Institute, Tullahoma, Tennessee 37388 and School of Mathematics, The Institute for Advanced Study, Princeton, New Jersey 08540

(Received 26 September 1989; accepted for publication 11 October 1989)

It is shown that supersymmetry can weaken integrability: A supersymmetric extension of the quasiclassical limit of the KP hierarchy has flows that no longer commute between themselves but still have an infinite common set of conserved densities. This extension has, thus, no Hamiltonian structure.

I. INTRODUCTION

The basic two-dimensional integrable system, the Benney system,[1] has the following moment representation:

$$A_{i,t} = A_{i+1,x} + iA_{i-1}A_{0,x}, \quad i \in \mathbb{Z}_+, \tag{1.1}$$

where $A_i = A_i(x,t)$ and $(\cdot\cdot)_{,(\cdot)} := \partial(\cdot\cdot)/\partial(\cdot)$. The Benney system (1.1) possesses many remarkable properties, in particular:

(i) There exists an infinite number of polynomial conserved densities[1]

$$\overline{H}_i \in A_i + \mathbb{Z}[A_0,\ldots,A_{i-2}], \quad i \in \mathbb{Z}_+, \tag{1.2}$$

starting with

$$\begin{aligned} &\overline{H}_0 = A_0, \\ &\overline{H}_1 = A_1, \\ &\overline{H}_2 = A_2 + A_0^2, \\ &\overline{H}_3 = A_3 + 3A_0A_1, \\ &\overline{H}_4 = A_4 + 4A_0A_2 + 2A_1^2 + 2A_0^3, \\ &\overline{H}_5 = A_5 + 5A_0A_3 + 5A_1A_2 + 10A_0^2A_1, \ldots . \end{aligned} \tag{1.3}$$

(ii) There exists an infinite number of "higher" Benney equations having the same infinite set (1.3) of polynomial conserved densities.[2,3] In particular, the next flow has the form[3]

$$A_{i,t} = A_{i+2,x} + A_0A_{i,x} + (i+1)A_iA_{0,x} + iA_{i-1}A_{1,x}, \quad i \in \mathbb{Z}_+. \tag{1.4}$$

(iii) All these flows commute between themselves.[3]

(iv) All these flows are Hamiltonian, with the Hamiltonian structure[3]

$$B_{ij} = iA_{i+j-1}\partial + \partial jA_{i+j-1}, \quad i,j \in \mathbb{Z}_+, \qquad \partial := \partial/\partial x, \tag{1.5}$$

so that the flow $\#m$ can be written as

$$A_{i,t} = \sum_j B_{ij}\left(\frac{\partial H}{\partial A_j}\right), \quad H := \frac{1}{m}\overline{H}_m, \quad m \in \mathrm{N}. \tag{1.6}$$

(v) All these flows have a common Poisson representation[4]:

$$L_{,t} = \{P_+, L\} = \{L, P_-\}, \tag{1.7}$$

where

$$L = \xi + \sum_{i=0}^{\infty} A_i \xi^{-i-1}. \tag{1.8}$$

Here P is an element of the Poisson centralizer $Z(L)$ of L in the ring $\overline{\mathscr{O}} := \overline{\mathscr{A}}((\xi^{-1}))$, $\overline{\mathscr{A}}$ being the minimal differential Q-algebra generated by ∂ and the A_i's:

$$\overline{\mathscr{A}} = \mathrm{Q}[A_i^{(j)}], \quad i,j \in \mathbb{Z}_+, \tag{1.9}$$

with a derivation ∂ acting on the polynomial generators of $\overline{\mathscr{A}}$ by the standard rule[5]

$$\partial(A_i^{(j)}) = A_i^{(j+1)}. \tag{1.10}$$

Thus, $Z(L)$ is generated over Q by $\{L^{\ell} | \ell \in \mathbb{Z}\}$; for an element

$$\sum_l p_l \xi^l \in \overline{\mathscr{O}},$$

we define

$$\left(\sum p_l\xi^l\right)_+ := \sum_{l \geq 0} p_l \xi^l, \quad \left(\sum p_l\xi^l\right)_- := \sum_{l<0} p_l\xi^l; \tag{1.11}$$

finally, the Poisson bracket {,} figuring in formula (1.7) is the standard one on $T^*(\mathbb{R}^1)$:

$$\{F,G\} := F_{,\xi}G_{,x} - F_{,x}G_{,\xi}. \tag{1.12}$$

In particular, the flow $\#m$ (1.6) has the Poisson representation (1.7) with $P = (1/m)L^m$; the flows (1.1) and (1.4) are the flows $\#2$ and $\#3$, respectively.

The properties (i)–(v) are not logically independent. For example, the flows commute (iii) since they are Hamiltonian (iv) and all the Hamiltonians are in involution (ii). But more importantly, the properties (i)–(iv), and many others, all follow from the single Poisson representation property (v), even when L in (1.8) is taken to be of the general form (Remark 2.28 in Ref. 6):

$$\mathscr{L} = \xi^M + \sum_{l=Q}^{M-2} u_l \xi^l, \quad M \in N, \tag{1.13}$$

$$Q = 0 \text{ or } Q = -\infty. \tag{1.14}$$

The purpose of this paper is to examine what happens with the flows (1.7) when one extends the plane $T^*(\mathbb{R}^1) = \mathbb{R}^2$ into the superplane $\mathbb{R}^{2|N}$ equipped with the (super)-Poisson bracket

$$\{F,G\} := F_{,\xi}G_{,x} - F_{,x}G_{,\xi} + \frac{1}{2\xi}\sum_{r=1}^{N} \mathscr{D}_r(F)\mathscr{D}_r(G), \tag{1.15}$$

F and G are even, where

 0022-2488/90/020459-11$03.00

$$\mathscr{D}_r := \frac{\partial}{\partial \theta_r} + \theta_r \frac{\partial}{\partial x}, \quad 1 \leqslant r \leqslant N, \tag{1.16}$$

are odd supercommuting derivations satisfying

$$[\mathscr{D}_r, \mathscr{D}_{\bar{r}}] := \mathscr{D}_r \mathscr{D}_{\bar{r}} + \mathscr{D}_{\bar{r}} \mathscr{D}_r = 2\delta_{r\bar{r}} \partial, \tag{1.17}$$

$\theta_1,...,\theta_N$ being the generators of the Grassmann algebra $\Lambda(N)$. Briefly summarizing the results of this examination, supersymmetry destroys integrability, but not entirely. Namely, we shall prove that the superextended flows *do not* commute between themselves, but nevertheless, all these flows *do have a common infinite set* of polynomial conserved densities. It follows that the new hierarchy of flows *is not Hamiltonian* (if it were, the flows would have commuted between themselves which they do not), despite having a *common* infinite set of conserved densities. This type of situation has been unknown so far in the theories of integrable and superintegrable systems; the term *semi-integrable* describes a system of noncommuting flows with a *common* set of conserved densities (or integrals, in the zero-dimensional case of classical mechanics).

At the moment, let me explain what the sources of troubles are. Firstly, the commutativity of the flows (1.7) can be traced by Wilson's[7] type of arguments, to the relation

$$\{\bar{\mathscr{O}}_+, \bar{\mathscr{O}}_+\} \subset \bar{\mathscr{O}}_+, \tag{1.18}$$

where

$$\bar{\mathscr{O}}_+ := \left\{ \sum_{l \geqslant 0} p_l \xi^l \,|\, p_l \in \mathscr{A} \right\}. \tag{1.19}$$

For the super-Poisson bracket (1.15), the relation (1.18) is no longer true. Secondly, in the classical even case, the existence of the conserved densities (1.3) for the flows (1.7) is predicated upon the existence of a residue for the Poissson bracket (1.12):

$$\operatorname{Res}\left(\sum p_l \xi^l\right) := \operatorname{res}_{-1}\left(\sum p_l \xi^l\right), \tag{1.20}$$

where

$$\operatorname{res}_\gamma\left(\sum p_l \xi^l\right) := p_\gamma, \quad \gamma \in \mathbb{Z}, \tag{1.21}$$

with the characteristic property

$$\operatorname{Res}(\{\bar{\mathscr{O}}, \bar{\mathscr{O}}\}) \sim 0; \tag{1.22}$$

here $a_1 \sim a_2$ means: $(a_1 - a_2) \in \operatorname{Im} \partial$ in the even case $a_1, a_2 \in \mathscr{A}$, and $(a_1 - a_2) \in \operatorname{Im} \partial + \Sigma_r \operatorname{Im} \mathscr{D}_r$ in the supercase $a_1, a_2 \in \mathscr{A}$, where $\mathscr{A}$ is the minimal differential commutative superalgebra over $\mathbb{Q}$ generated by the ∂, $\mathscr{D}_r$'s, and A_i's. However, for the super-Poisson bracket (1.15), such a residue does not exist when N is an odd integer. This can be seen from the following calculation: for even $a,b \in \mathscr{A}$,

$$\{a\xi^l, b\xi^\gamma\} \text{ [by (1.15)]}$$
$$= \xi^{l+\gamma-1}[la\partial(b) - \gamma\partial(a)b + \tfrac{1}{2}\mathscr{D}_r(a)\mathscr{D}_r(b)]$$

(the repeated index r is always summed between 1 and N), so that

$$\operatorname{res}_{l+\gamma-1}(\{a\xi^l, b\xi^\gamma\})$$
$$= la\partial(b) - \gamma\partial(a)b + \tfrac{1}{2}\mathscr{D}_r(a)\mathscr{D}_r(b).$$

But, $\partial(a)b \sim -a\partial(b)$ and

$$\tfrac{1}{2}\mathscr{D}_r(a)\mathscr{D}_r(b)$$
$$= \tfrac{1}{2}\mathscr{D}_r[a\mathscr{D}_r(b)] - \tfrac{1}{2}a\mathscr{D}_r\mathscr{D}_r(b) \text{ [by (1.17)]}$$
$$\sim -(N/2)a\partial(b), \tag{1.23}$$

so that

$$\operatorname{res}(\{a\xi^\ell, b\xi^\gamma\}) \sim (\ell + \gamma - N/2)a\partial(b). \tag{1.24}$$

Hence, for $\mathscr{O} := \{\Sigma p_l \xi^l \,|\, p_l \in \mathscr{A}\}$, $\operatorname{res}_?(\{\mathscr{O}, \mathscr{O}\}) \sim 0$ iff

$$? + 1 - N/2 = 0,$$

i.e.,

$$\operatorname{Res} := \operatorname{res}_{(N/2)-1}, \tag{1.25}$$

which makes sense only when $N/2 \in \mathbb{Z}_+$. Thus, for $N \in 1 + 2\mathbb{Z}_+$, there is no reason why any one of the flows should have an infinity of conserved densities, and even less why this infinity of conserved densities should be the *same* for each of the flows. No doubt such a reason exists, but I was unable to find it. For a nonexpert reader, here is a quick construction of an infinity of conserved densities for the case $N \in 2\mathbb{Z}_+$, when a residue does exist: From the equation

$$\mathscr{L}_{,t} = \{\mathscr{P}_+, \mathscr{L}\}, \quad \forall\, \mathscr{P} \in Z(\mathscr{L}),$$

we get, since $\{\cdot,\cdot\}$ is a derivation with respect to each argument,

$$(\mathscr{L}^{m/M})_{,t} = \{\mathscr{P}_+, \mathscr{L}^{m/M}\}, \quad \forall\, m \in \mathbb{N},$$

so that

$$[\operatorname{Res}(\mathscr{L}^{m/M})]_{,t} = \operatorname{Res}[(\mathscr{L}^{m/M})_{,t}]$$
$$= \operatorname{Res}(\{\mathscr{P}_+, \mathscr{L}^{m/M}\}) \sim 0. \tag{1.26}$$

The paper is organized as follows. In the next section we prove the noncommutativity of the flows. Section III is devoted to combinatorial properties of the polynomials $\bar{H}_i$'s (1.3); these properties are an ingredient in the proof of existence of an infinity of conserved densities for the case of arbitrary N, Sec. IV.

I conclude this Introduction with a few remarks on the formula (1.15) for the super-Poisson bracket. One can directly check the Jacobi identity for this bracket; to avoid the lengthy check, one can notice instead that the bracket (1.15) is associated in the standard way with the nondegenerate differential two-form

$$\omega = (dx - \theta_r\, d\theta_r) \wedge d\xi + \xi\, d\theta_r \wedge d\theta_r, \tag{1.27}$$

so that

$$X_F \lrcorner\, \omega = d(F), \quad \forall F, \tag{1.28}$$

for

$$X_F = F_{,\xi}\frac{\partial}{\partial x} - F_{,x}\frac{\partial}{\partial \xi} + \frac{1}{2\xi}\mathscr{D}_r(F)\mathscr{D}_r, \tag{1.29}$$

and

$$\{F,G\} = X_F(G).$$

Obviously, the two-form ω is closed, and even exact: $\omega = -d[\xi(dx - \theta_r\, d\theta_r)]$. [In addition, the Lie derivative of ω vanishes along the derivations $\bar{\mathscr{D}}_r := \partial/\partial\theta_r - \theta_r(\partial/\partial x)$, since $\bar{\mathscr{D}}_r \lrcorner\, \omega = d(2\xi\, d\theta_r)$.] It follows that one has a multidimensional analog of the Poisson bracket (1.15):

$$\{F,G\} = \sum_x \left(\frac{\partial F}{\partial \xi_x} \frac{\partial G}{\partial x_x} - \frac{\partial F}{\partial x_x} \frac{\partial G}{\partial \xi_x} + \frac{1}{2\xi_x} \sum_{r=1}^{N(x)} \mathscr{D}_{r|x}(F) \mathscr{D}_{r|x}(G) \right), \quad (1.30)$$

where

$$\mathscr{D}_{r|x} := \frac{\partial}{\partial \theta_{r|x}} + \theta_{r|x} \frac{\partial}{\partial x_x};$$

the associated symplectic form is simply

$$\omega = \sum_x [(dx_x - \theta_{r|x} \, d\theta_{r|x}) \wedge d\xi_x + \xi_x \, d\theta_{r|x} \wedge d\theta_{r|x}]. \quad (1.31)$$

Changing ξ in (1.15) into ξ^λ and then factoring ξ^λ out of F and G results in the Poisson bracket

$$\{F,G\} = FG_{,x} - F_{,x}G + \tfrac{1}{2} \mathscr{D}_r(F) \mathscr{D}_r(G) \quad (1.32a)$$

$$+ (1/\lambda)\xi(F_{,\xi}G_{,x} - F_{,x}G_{,\xi}). \quad (1.32b)$$

Since λ is an arbitrary constant, an arbitrary linear combination of the expressions (1.32a) and (1.32b) with constant coefficients is also a Poisson bracket:

$$\{F,G\} = \beta [FG_{,x} - F_{,x}G + \tfrac{1}{2} \mathscr{D}_r(F) \mathscr{D}_r(G)] + \alpha\xi(F_{,\xi}G_{,x} - F_{,x}G_{,\xi}). \quad (1.33)$$

If we consider F and G polynomial in ξ, the Poisson bracket (1.33) induces the Lie algebra structure on the space of infinite columns whose entries are functions of (x,θ):

$$[X,Y]_n = \sum_{i+j=n} \{[(\alpha i + \beta)X_i Y_{j,x} - (\alpha j + \beta) X_{i,x} Y_j + (\beta/2) \mathscr{D}_r(X_i) \mathscr{D}_r(Y_j)]\}, \quad i,j,n \in \mathbb{Z}_+. \quad (1.34)$$

The associated Hamiltonian matrix B, computed from the defining relation[8]

$$B(X)^t Y \sim -\sum_n A_n [X,Y]_n, \quad (1.35)$$

is

$$B_{ij} = (\alpha i + \beta) A_{i+j} \partial + \partial(\alpha j + \beta) A_{i+j} - (\beta/2) \mathscr{D}_r A_{i+j} \mathscr{D}_r. \quad (1.36)$$

It can be shown by a direct calculation that the Hamiltonian matrix (1.36) is the only Hamiltonian matrix of the form

$$B_{ij} = (\alpha i + \beta) A_{i+j} \partial + \partial(\alpha j + \beta) A_{i+j} - \epsilon \mathscr{D}_r A_{i+j} \mathscr{D}_r, \quad (1.37)$$

in addition to the purely even solution $\{\epsilon = 0\}$.[9] Notice that when F and G are ξ-independent, the Poisson bracket (1.32) reduces to the Poisson bracket (1.32a) that defines the Lie algebra of functions on a N-(super)circle; the same bracket arises out of the bracket (1.15) when F and G are linear in ξ.

II. SUPERFLOWS

Let

$$\mathscr{L} = \xi^M + \sum_{l=-\infty}^{M-2} u_l \xi^l, \quad M \in \mathbb{N},$$

$$u_l = u_l(x,\theta,t) \text{ is even } \forall l, \quad (2.1)$$

and let

$$\mathscr{P} = \mathscr{L}^{m/M} = \xi + ..., \quad m \in \mathbb{N}, \quad (2.2)$$

be a $\mathbb{Q}$-generator of positive ξ-degree of the Poisson centralizer $Z(L)$ of L in the ring $\mathscr{O} := \mathscr{A}((\xi^{-1}))$, where now $\mathscr{A}$ is generated over $\mathbb{Q}$ by the ∂, $\mathscr{D}_r$'s, and u_l's. We consider an evolutionary derivation $\partial_{;\mathscr{P}}$ of $\mathscr{A}$ (i.e., commuting with the actions of $\mathbb{Q}$, ∂, and $\mathscr{D}_r$'s), defined by the rule

$$\partial_{;\mathscr{P}}(\mathscr{L}) = \{\mathscr{P}_+, \mathscr{L}\} \quad (2.3a)$$

$$= \{\mathscr{L}, \mathscr{P}_-\}, \quad (2.3b)$$

with the Poisson bracket $\{,\}$ defined by formula (1.15), and with the usual understanding that $\partial_{;\mathscr{P}}$ (like other derivations ∂_t, ∂, $\mathscr{D}_r$'s) acts trivially on ξ. Thus, the action of $\partial_{;\mathscr{P}}$ on $\mathscr{A}$ can be read off formula (2.3):

$$\partial_{;\mathscr{P}}(u_l) = \text{res}_l[\partial_{;\mathscr{P}}(\mathscr{L})] = \text{res}_l(\{\mathscr{P}_+, \mathscr{L}\}) = \text{res}_l(\{\mathscr{L}, \mathscr{P}_-\}). \quad (2.4)$$

Note that the expressions (2.3a) and (2.3b) agree between themselves since

$$0 = \{\mathscr{P}, \mathscr{L}\} = \{\mathscr{P}_+ + \mathscr{P}_-, \mathscr{L}\} = \{\mathscr{P}_+, \mathscr{L}\} - \{\mathscr{L}, \mathscr{P}_-\}. \quad (2.5)$$

It remains to verify that the derivation $\partial_{;\mathscr{P}}$ is correctly defined: from the form of $\mathscr{L}$ (2.1) we see that $\partial_{;\mathscr{P}}(\mathscr{L})$ must belong to $\mathscr{O}_{<M-2}$, where

$$\mathscr{O}_{<\gamma} := \left\{ \sum_{l<\gamma} p_l \xi^l \,|\, p_l \in \mathscr{A} \right\}. \quad (2.6)$$

But, by formula (1.15),

$$\{\mathscr{O}_{<\gamma}, \mathscr{O}_{<l}\} \subset \mathscr{O}_{<\gamma+l-1}. \quad (2.7)$$

Hence, by formula (2.3b)

$$\partial_{;\mathscr{P}}(\mathscr{L}) = \{\mathscr{L}, \mathscr{P}_-\} \in \{\mathscr{O}_{<M}, \mathscr{O}_{<-1}\} \subset \mathscr{O}_{<M-2}, \quad (2.8)$$

as desired. In case the reader is wondering why $\mathscr{P}$ has been taken of positive ξ-degree, formula (2.3a) shows that a negative ξ-degree $\mathscr{P}$ yields $\partial_{;\mathscr{P}} = 0$, and the same happens when the ξ-degree of $\mathscr{P}$ is zero: in this case $\mathscr{P} = \mathscr{L}^0 = 1$.

Suppose now that $\mathscr{R}$ is another element of $Z(\mathscr{L})$ of positive ξ-degree:

$$\mathscr{R} = \mathscr{L}^{\bar{m}/M}, \quad \bar{m} \in \mathbb{N}, \bar{m} \neq m, \quad (2.9)$$

and let $\partial_{;\mathscr{R}}$ be the corresponding evolutionary derivation of $\mathscr{A}$, given by the formulas

$$\partial_{;\mathscr{R}}(\mathscr{L}) = \{\mathscr{R}_+, \mathscr{L}\} \quad (2.10a)$$

$$= \{\mathscr{L}, \mathscr{R}_-\}. \quad (2.10b)$$

We are going to show that, in contrast to the purely even case (when the $\mathscr{D}_r$'s are absent), the derivations $\partial_{;\mathscr{P}}$ and $\partial_{;\mathscr{R}}$ do not commute.

Theorem 2.11:

$$[\partial_{\mathscr{P}},\partial_{\mathscr{R}}](\mathscr{L})$$
$$=\{(2\xi)^{-1}\mathscr{D}_r[\mathrm{res}_0(\mathscr{R})]\mathscr{D}_r[\mathrm{res}_0(\mathscr{P})],\mathscr{L}\}. \tag{2.12}$$

Proof: Since the Poisson bracket (1.15) is a derivation with respect to each argument, formulas (2.3) imply that

$$\partial_{\mathscr{P}}(\mathscr{R})=\{\mathscr{P}_+,\mathscr{R}\} \tag{2.13a}$$
$$=\{\mathscr{R},\mathscr{P}_-\}. \tag{2.13b}$$

In particular,

$$\partial_{\mathscr{P}}(\mathscr{R}_+)=[\partial_{\mathscr{P}}(\mathscr{R})]_+[\text{by (2.13b)}]=\{\mathscr{R},\mathscr{P}_-\}_+. \tag{2.14}$$

Hence,

$$\partial_{\mathscr{P}}\partial_{\mathscr{R}}(\mathscr{L})\ [\text{by (2.10a)}]$$
$$=\partial_{\mathscr{P}}(\{\mathscr{R}_+,\mathscr{L}\})$$
$$[\text{since } \partial_{\mathscr{P}} \text{ commutes with the } \partial, \mathscr{D}_r\text{'s},\xi]$$
$$=\{\partial_{\mathscr{P}}(\mathscr{R}_+),\mathscr{L}\}+\{\mathscr{R}_+,\partial_{\mathscr{P}}(\mathscr{L})\}$$
$$[\text{by (2.14),(2.3a)}]$$
$$=\{\{\mathscr{R},\mathscr{P}_-\}_+,\mathscr{L}\}+\{\mathscr{R}_+,\{\mathscr{P}_+,\mathscr{L}\}\}. \tag{2.15}$$

Interchanging $\mathscr{P}$ and $\mathscr{R}$ in formula (2.15), we obtain

$$\partial_{\mathscr{R}}\partial_{\mathscr{P}}(\mathscr{L})=\{\{\mathscr{P},\mathscr{R}_-\}_+,\mathscr{L}\}+\{\mathscr{P}_+,\{\mathscr{R}_+,\mathscr{L}\}\}. \tag{2.16}$$

Subtracting formula (2.16) from formula (2.15) and using the Jacobi identity, we get

$$[\partial_{\mathscr{P}},\partial_{\mathscr{R}}](\mathscr{L})=\{\Delta,\mathscr{L}\}, \tag{2.17}$$

where

$$\Delta:=\{\mathscr{R},\mathscr{P}_-\}_++\{\mathscr{R}_-,\mathscr{P}\}_++\{\mathscr{R}_+,\mathscr{P}_+\}. \tag{2.18}$$

Now, by formula (1.15),

$$\{\mathscr{O}_{<0},\mathscr{O}_{<0}\}_+=\{0\}, \tag{2.19}$$

so that

$$\Delta=\{\mathscr{R}_+,\mathscr{P}_-\}_++\{\mathscr{R}_-,\mathscr{P}_+\}_++\{\mathscr{R}_+,\mathscr{P}_+\}_+ \tag{2.20a}$$
$$+\{\mathscr{R}_+,\mathscr{P}_+\}-\{\mathscr{R}_+,\mathscr{P}_+\}_+. \tag{2.20b}$$

But, by formulas (2.2) and (2.9),

$$\{\mathscr{R},\mathscr{P}\}=0. \tag{2.21}$$

Hence

$$0=\{\mathscr{R},\mathscr{P}\}_+=\{\mathscr{R}_++\mathscr{R}_-,\mathscr{P}_++\mathscr{P}_-\}_+\ [\text{by (2.19)}]$$
$$=\{\mathscr{R}_+,\mathscr{P}_-\}_++\{\mathscr{R}_-,\mathscr{P}_+\}_+$$
$$+\{\mathscr{R}_+,\mathscr{P}_+\}_+,$$

and formula (2.20) becomes

$$\Delta=\{\mathscr{R}_+,\mathscr{P}_+\}-\{\mathscr{R}_+,\mathscr{P}_+\}_+\ [\text{by definition of } \mathscr{O}_-]$$
$$=\{\mathscr{R}_+,\mathscr{P}_+\}_-\ [\text{by (1.15)}] \tag{2.22a}$$
$$=(2\xi)^{-1}\mathscr{D}_r[\mathrm{res}_0(\mathscr{R})]\mathscr{D}_r[\mathrm{res}_0(\mathscr{P})]. \tag{2.22b}$$

Substituting formula (2.22b) into formula (2.17), we get the desired formula (2.12). ■

Since, for $m\neq\bar{m}$, $\mathrm{res}_0(\mathscr{R})$ and $\mathrm{res}_0(\mathscr{P})$ are two polynomials in the entries u_r's not all of which are the same, the expression (2.22b) does not vanish unless one of the $\mathrm{res}_0(\mathscr{R})$ and $\mathrm{res}_0(\mathscr{P})$ does, which happens precisely when either m or $\bar{m}$ equals one. In such a case, say for $m=1$,

$$\partial_{\mathscr{P}}=\partial, \tag{2.23}$$

which obviously commutes with all the $\partial_{\mathscr{R}}$'s. In the general case, when neither m nor $\bar{m}$ equals one,

$$[\partial_{\mathscr{P}},\partial_{\mathscr{Q}}]\neq 0. \tag{2.24}$$

Remark 2.25: The map

$$\mathscr{L}\mapsto L=\mathscr{L}^{1/M}=\xi+\sum_{i=0}^{\infty}A_i\xi^{-i-1} \tag{2.26}$$

takes the derivation $\partial_{\mathscr{P}}$ (2.3) into the derivation ∂_P (1.7); the inverse of this map, $L\mapsto\mathscr{L}=L^M$, takes ∂_P into $\partial_{\mathscr{P}}$. This map is, thus, an isomorphism. From now on, therefore, we can and shall work with L only.

Remark 2.27: The classical Poisson bracket (1.12) on $T^*(\mathbb{R}^1)$ is the quasiclassical (= zero dispersion) limit of the commutator

$$[F,G]=F\circ G-G\circ F, \tag{2.28}$$

where $\circ$ is the multiplication in the associative ring of pseudodifferential operators,[5]

$$F\circ G:=\sum_{n\geq 0}\frac{1}{n!}\frac{\partial^n F}{\partial\xi^n}\partial^n(G). \tag{2.29}$$

Consequently, the Benney hierarchy is the quasiclassical limit of the *KP* hierarchy. Since the $N=1$ supersymmetric *KP* hierarchy of Manin and Radul[10] does *not* have the quasiclassical limit, the question arises whether the super-Poisson bracket (1.15) is the first nontrivial term of the commutator resulting from an associative product that extends by the $\mathscr{D}_r$'s the usual multiplication (2.29). If such an extension exists, it would provide a new supersymmetric extension of the *KP* hierarchy, and for *arbitrary N.*

III. COMBINATORIAL FORMULAS

In this section we derive various explicit formulas concerning the Benney hierarchy and its superextensions.

We start with the grading

$$rk(A_i)=i+2,\quad rk(\xi)=1,\quad rk(\partial)=1,\quad rk(\mathscr{D}_r)=\tfrac{1}{2},$$
$$rk(\mathrm{Q})=0. \tag{3.1}$$

In this grading,

$$rk(L)=1, \tag{3.2}$$

so that

$$rk(\mathrm{res}_{-1}(L^m))=m+1. \tag{3.3}$$

Let

$$\bar{H}_m:=\mathrm{res}_{-1}(L^{m+1})/(m+1), \tag{3.4}$$

so that, by (3.3),

$$rk(\bar{H}_m)=m+2. \tag{3.5}$$

Theorem 3.6: The homogeneous polynomials $\bar{H}_m$ of rank $m+2$ are uniquely defined by the formulas

$$\left.\frac{\partial^k\bar{H}_m}{\partial_{A_{i_1}}\cdots\partial A_{i_k}}\right|_{A=0}=\frac{m!}{(m-k+1)!}\delta\Big(m+2,\sum_s(i_s+2)\Big), \tag{3.7}$$

where

$$\delta(\cdot,\cdot\cdot):=\delta^{\cdot}_{\cdot\cdot}. \tag{3.8}$$

Proof: The uniqueness is obvious. To prove formula (3.7), we note that, by (1.8),

$$\frac{\partial}{\partial A_{i_1}}(L^{m+1})=(m+1)L^m\xi^{-(i_1+1)},$$

so that

$$\frac{\partial}{\partial A_{i_1}}\cdots\frac{\partial}{\partial A_{i_k}}(L^{m+1})$$
$$=(m+1)\cdots(m+1-k+1)L^{m+1-k}\xi^{-\Sigma(i_s+1)}, \tag{3.9}$$

whence

$$\left.\frac{\partial^k\overline{H}_m}{\partial A_{i_1}\cdots\partial A_{i_k}}\right|_{A=0} \quad [\text{by } (3.4)]$$
$$=\frac{1}{m+1}\operatorname{res}_{-1}\left(\left[\frac{\partial}{\partial A_{i_1}}\cdots\frac{\partial}{\partial A_{i_k}}(L^{m+1})\right]\bigg|_{A=0}\right)$$
$$[\text{by } (3.9)]$$
$$=\frac{1}{m+1}\operatorname{res}_{-1}\left(\frac{(m+1)!}{(m+1-k)!}\xi^{m+1-k-\Sigma(i_s+1)}\right)$$
$$=\frac{m!}{(m+1-k)!}\delta\Big(m+1-\sum(i_s+2),-1\Big),$$

which is the same as (3.7).

The same proof as above shows that

$$\left(\frac{\partial^k}{\partial A_{i_1}\cdots\partial A_{i_k}}\left[\operatorname{res}_\ell\left(\frac{L^{m+1}}{m+1}\right)\right]\right)\bigg|_{A=0}$$
$$=\frac{m!}{(m+1-k)!}\delta\Big(m+1-\ell,\sum(i_s+2)\Big). \tag{3.10}$$

Hence, fixing $\ell\in\mathbb{Z}_+\cup\{-1\}$, and defining

$$H_m(l):=\operatorname{res}_l(L^{m+l+2})/(m+l+2), \tag{3.11}$$

so that

$$H_m(-1)=\overline{H}_m, \tag{3.12}$$

and

$$rk(H_m(l))=m+2, \tag{3.13}$$

we get

$$\left.\frac{\partial^k H_m(l)}{\partial A_{i_1}\cdots\partial A_{i_k}}\right|_{A=0}$$
$$=\left[\frac{\partial^k}{\partial A_{i_1}\cdots\partial A_{i_k}}\left(\frac{\operatorname{res}_l(L^{m+l+2})}{m+l+2}\right)\right]\bigg|_{A=0}$$
$$[\text{by } (3.10)]$$
$$=\frac{(m+l+1)!}{(m+l+2-k)!}\delta\Big(m+2,\sum(i_s+2)\Big). \tag{3.14}$$

In particular, when $N\in 2\mathbb{Z}_+$ and $l=N/2-1$ (1.24), we obtain

$$\left[\frac{\partial^k}{\partial A_{i_1}\cdots\partial A_{i_k}}(H_m(N/2-1))\right]\bigg|_{A=0}$$
$$=\frac{(m+N/2)!}{(m+N/2-k+1)!}\delta\Big(m+2,\sum(i_s+2)\Big) \tag{3.15a}$$
$$=\frac{1}{m+N/2-k+1}(m+N/2)\cdots$$
$$(m+N/2-k+1)\delta\Big(m+2,\sum(i_s+2)\Big). \tag{3.15b}$$

Thus, for even N, formula (3.15a) provides us with an explicit formula for the infinity of polynomial conserved densities for the superextended hierarchy (1.7), (1.15); however, formula (3.15b) makes sense even when N is odd, and it is natural to suppose that this formula provides the desired conserved densities for *all* values of N. Even if this supposition is true (it is), it is difficult to verify directly. Let us see why.

For $m=2$ and $P=L^2/2$, we have

$$\tfrac{1}{2}(L^2)_+=\tfrac{1}{2}(\xi^2+2A_0)=\xi^2/2+A_0, \tag{3.16}$$

so that the motion equations (1.7) with the Poisson bracket (1.15) are

$$A_{i,t}=A_{i+1,x}+iA_{i-1}A_{0,x}+\tfrac{1}{2}\mathscr{D}_r(A_0)\mathscr{D}_r(A_{i-1}), \tag{3.17}$$

cf. (1.1). Similarly, for $m=3$ and $P=L^3/3$, we get

$$\tfrac{1}{3}(L^3)_+=\tfrac{1}{3}(\xi^3+3A_0\xi+3A_1)=\xi^3/3+A_0\xi+A_1, \tag{3.18}$$

so that the corresponding motion equations (1.7) are

$$A_{i,t}=A_{i+2,x}+A_0A_{i,x}+(i+1)A_iA_{0,x}$$
$$+iA_{i-1}A_{1,x} \tag{3.19a}$$
$$+\tfrac{1}{2}\mathscr{D}_r(A_0)\mathscr{D}_r(A_i)+\tfrac{1}{2}\mathscr{D}_r(A_1)\mathscr{D}_r(A_{i-1}), \tag{3.19b}$$

cf. (1.4). The super-Benney system (3.17) can be directly checked to have the following conserved densities:

$$\begin{aligned}
H_0&=A_0,\\
H_1&=A_1,\\
H_2&=A_2+(2+N/2)(A_0^2/2),\\
H_3&=A_3+(3+N/2)A_0A_1,\\
H_4&=A_4+\left(4+\frac{N}{2}\right)\left(A_0A_2+\frac{A_1^2}{2}\right)\\
&\quad+\left(4+\frac{N}{2}\right)\left(3+\frac{N}{2}\right)\frac{A_0^3}{6},\\
H_5&=A_5+\left(5+\frac{N}{2}\right)(A_0A_3+A_1A_2)\\
&\quad+\left(5+\frac{N}{2}\right)\left(4+\frac{N}{2}\right)\frac{A_0^2A_1}{2},\ldots,
\end{aligned} \tag{3.20}$$

cf. (1.3). We see that formulas (3.20) agree with the formula (3.15b). To show that the list (3.20) can be continued indefinitely, define the polynomials

$$H_m \in \mathbb{Q}[A_0,\dots,A_m], \quad m \in \mathbb{Z}_+, \tag{3.21}$$

by the formulas

$$\left.\frac{\partial^k H_m}{\partial A_{i_1}\cdots\partial A_{i_k}}\right|_{A=0} = \delta\Big(m+2, \sum (i_s+2)\Big) c_{m|k}, \tag{3.22}$$

where the constants $c_{m|k}$, appearing in formula (3.15b), are defined by the rule

$$c_{m|1} = 1, \quad c_{m|k+1} = c_{m|k}(m + N/2 - k + 1). \tag{3.23}$$

Theorem 3.24: Define the homogeneous polynomials

$$F_m \in \mathbb{Q}[A_0,\dots,A_{m+1}], \quad rk(F_m) = m+3, \tag{3.25}$$

$$\widetilde{F}_m \in \mathbb{Q}[A_0,\dots,A_{m-1}], \quad \widetilde{F}_0 = 0, \quad rk(\widetilde{F}_m) = m+1, \tag{3.26}$$

by the formulas

$$\left.\frac{\partial^k F_m}{\partial A_{i_1}\cdots\partial A_{i_k}}\right|_{A=0} = \delta\Big(m+3, \sum (i_s+2)\Big) c_{m|k}, \tag{3.27}$$

$$\left.\frac{\partial^k \widetilde{F}_m}{\partial A_{i_1}\cdots\partial A_{i_k}}\right|_{A=0} = -\frac{1}{2}\delta\Big(m+1, \sum (i_s+2)\Big) c_{m|k}. \tag{3.28}$$

Then the motion equations (3.17) imply

$$H_{m,t} = F_{m,x} + \mathscr{D}_r[\widetilde{F}_m \mathscr{D}_r(A_0)]. \tag{3.29}$$

(Thus, the H_m's are conserved densities of the system (3.17) for *any N*.)

Proof: Denote

$$(\cdot)_{,i} := \frac{\partial(\cdot)}{\partial A_i}. \tag{3.30}$$

Then, for the lhs of (3.29) we get

$$\begin{aligned} H_{m,t}\ &[\text{by } (3.17)] \\ &= H_{m,i}[A_{i+1,x} + iA_{i-1}A_{0,x} \\ &\quad + \tfrac{1}{2}\mathscr{D}_r(A_0)\mathscr{D}_r(A_{i-1})], \end{aligned} \tag{3.31a}$$

while for the rhs of (3.29) we obtain

$$\begin{aligned} &F_{m,x} + \mathscr{D}_r[\widetilde{F}_m\mathscr{D}_r(A_0)] \\ &\quad = F_{m,i}A_{i,x} + N\widetilde{F}_m A_{0,x} + \widetilde{F}_{m,i}\mathscr{D}_r(A_i)\mathscr{D}_r(A_0). \end{aligned} \tag{3.31b}$$

Comparing (3.31a) with (3.31b) we see that they are equal provided they have the same coefficients in front of $A_{i+1,x}$, $A_{0,x}$, and $\mathscr{D}_r(A_0)\mathscr{D}_r(A_{i+1})$, i.e., when

$$H_{m,i} = F_{m,i+1}, \tag{3.32}$$

$$iA_{i-1}H_{m,i} = F_{m,0} + N\widetilde{F}_m, \tag{3.33}$$

$$\tfrac{1}{2}H_{m,i+2} = -\widetilde{F}_{m,i+1}. \tag{3.34}$$

To prove formulas (3.32)–(3.34) we apply to each one of them the operator $\partial^k/(\partial A_{i_1}\cdots\partial A_{i_k})|_{A=0}$. Then formulas (3.32) and (3.34), in view of formulas (3.22), (3.27), and (3.28), turn at once into identities, while the constant terms are 1 for $i=m$ and $\frac{1}{2}$ for $i=m-2$ in (3.32) and (3.34), respectively. With the formula (3.33) we need a bit of extra work. We have, for the lhs:

$$\begin{aligned} &\frac{\partial^k}{\partial A_{i_1}\cdots\partial A_{i_k}}(iA_{i-1}H_{m,i})\Big|_{A=0} \\ &= \sum_{s=1}^{k}(i_s+1)\frac{\partial^{k-1}}{\partial A_{i_1}\cdots\widehat{\partial A_{i_s}}\cdots\partial A_{i_k}}\left(\frac{\partial H_m}{\partial A_{i_s+1}}\right)\Bigg|_{A=0} \\ &\quad [\text{by } (3.22)] \\ &= \sum (i_s+1)c_{m|k}\delta\Big(m+2, \sum(i_s+2)+1\Big) \\ &= (m+1-k)c_{m|k}\delta\Big(m+1, \sum(i_s+2)\Big), \end{aligned} \tag{3.35a}$$

and for the rhs:

$$\begin{aligned} &\frac{\partial^k}{\partial A_{i_1}\cdots\partial A_{i_k}}(F_{m,0} + N\widetilde{F}_m)|_{A=0}\ [\text{by } (3.27),(3.28)] \\ &= \delta\Big(m+3, \sum(i_s+2)+2\Big)c_{m|k+1} \\ &\quad + N(-\tfrac{1}{2})\delta\Big(m+1, \sum(i_s+2)\Big)c_{m|k} \\ &= \delta\Big(m+1, \sum(i_s+2)\Big)(c_{m|k+1} - (N/2)c_{m|k}). \end{aligned} \tag{3.35b}$$

Comparing expressions (3.35a) and (3.35b) we find that they are equal in view of formula (3.23). ■

If we now consider the time derivative of the polynomials H_m's along the flow #3 (3.19), we get

$$\begin{aligned} A_{0,t} &= (A_2 + A_0^2)_{,x}, \\ A_{1,t} &= (A_3 + 2A_0A_1)_{,x}, \\ [A_2 &+ (2+N/2)(A_0^2/2)]_{,t} \\ &= [A_4 + (3+N/2)A_0A_2 + A_1^2]_{,x} \\ &\quad + \mathscr{D}_r[-\tfrac{1}{2}A_2\mathscr{D}_r(A_0)]. \end{aligned} \tag{3.36}$$

Clearly, it is quite difficult to guess the structure of the fluxes here; so we cannot rely on the exact formulas of the type (3.29). We shall use a different procedure in the next section.

Remark 3.37: The flow (3.19) has an invariant submanifold

$$\{A_1 = A_3 = A_5 = \cdots = 0\}, \tag{3.38}$$

on which the remaining variables

$$R_n := A_{2n} \tag{3.39}$$

satisfy the equation

$$\begin{aligned} R_{n,t} &= R_{n+1,x} + R_0R_{n,x} + (2n+1)R_nR_{0,x} \\ &\quad + \tfrac{1}{2}\mathscr{D}_r(R_0)\mathscr{D}_r(R_n). \end{aligned} \tag{3.40}$$

Similar to the purely even case,[9] one can show that the following more general system:

$$\begin{aligned} R_{n,t} &= R_{n+1,x} + R_0R_{n,x} + (\alpha n + \beta)R_nR_{0,x} \\ &\quad + \epsilon\mathscr{D}_r(R_0)\mathscr{D}_r(R_n), \end{aligned} \tag{3.41}$$

where α, β, and ϵ are arbitrary (even) constants, has an infinite number of polynomial conserved densities

$$h_m \in \mathbb{Q}[R_0, \ldots, R_m; \alpha, \beta, \epsilon], \tag{3.42}$$

defined by the formulas

$$\left. \frac{\partial^k h_m}{\partial R_{i_1} \cdots \partial R_{i_k}} \right|_{R=0} = \delta\left(m+1, \sum (i_s + 1)\right) d_{m|k}, \tag{3.43}$$

where

$$d_{m|1} = 1,$$
$$d_{m|k+1} = d_{m|k}[(m-k+1)\alpha + \kappa\beta - 1 + N\epsilon]. \tag{3.44}$$

IV. CONSERVED DENSITIES

In this section we construct an infinity of polynomial conserved densities for the hierarchy (1.7), (1.15), and then show that these conserved densities are given by the explicit formulas (3.22), (3.23).

First we rewrite in components the motion equations (1.7), (1.15). Set

$$L^\gamma = \sum_l p_l(\gamma)\xi^l, \quad \forall \gamma \in \mathbb{Z}. \tag{4.1}$$

Then

$$d(\overline{H}_m) \text{ [by (3.4)]} = \text{res}_{-1}[L^m d(L)] \text{ [by (4.1)]}$$
$$= \text{res}_{-1}\left(\sum_l p_l(m)\xi^l \sum_i dA_i \xi^{-i-1}\right)$$
$$= \sum_i p_i(m) dA_i. \tag{4.2}$$

Hence,

$$p_i(m) = \frac{\partial \overline{H}_m}{\partial A_i}, \quad i \in \mathbb{Z}_+. \tag{4.3}$$

Now, since $L_{,t} \in \mathscr{O}_-$, the motion equations (1.7) can be written in the form

$$L_{,t} = \{P_+, L\}_-, \tag{4.4}$$

so that, for $P = L^m$, we obtain

$$L_{,t} = \sum A_{i,t}\xi^{-i-1}$$
$$= \left\{ \sum_{n>0} p_n(m)\xi^n, \xi + \sum A_i \xi^{-i-1} \right\}_- \quad \text{[by (1.15)]}$$
$$= \left\{ \sum_{n>0} p_n(m)\xi^n, \sum A_i \xi^{-i-1} \right\}_-$$

[by (1.15), and suppressing index m from $p_n(m)$]

$$= \left(\sum_{n,i>0} \xi^{n-i-2}[np_n A_{i,x} + (i+1)p_{n,x}A_i + \tfrac{1}{2}\mathscr{D}_r(p_n)\mathscr{D}_r(A_i)] \right)_-.$$

Hence, for $P = L^m$, the motion equations (1.7), (1.15) are:

$$A_{i,t} = \sum_{n>0} [nA_{n+i-1,x}p_n + (n+i)A_{n+i-1}p_{n,x} + \tfrac{1}{2}\mathscr{D}_r(p_n)\mathscr{D}_r(A_{n+i-1})], \quad i \in \mathbb{Z}_+, \tag{4.5}$$

where we remember that

$$p_n := p_n(m) = \frac{\partial \overline{H}_m}{\partial A_n}, \quad n \in \mathbb{Z}_+. \tag{4.6}$$

Let us now introduce the generating function

$$f = f(z) := \sum_{i=0}^{\infty} A_i z^i, \tag{4.7}$$

and let us convert the infinite system of motion equations (4.5) into a single motion equation for f. We have

$$f_{,t} \text{ [by (4.7)]} = \sum_i A_{i,t} z^i \text{ [by (4.5)]}$$
$$= \sum_{i,n} z^i [nA_{n+i-1,x}p_n + (n+i)A_{n+i-1}p_{n,x} + \tfrac{1}{2}\mathscr{D}_r(p_n)\mathscr{D}_r(A_{n+i-1})]. \tag{4.8}$$

We transform separately each of the three summands in the expression (4.8). We have,

$$\sum_{i,n} z^i n A_{n+i-1,x} p_n = \sum_n np_n z^{1-n}\left(\sum_i A_{n+i-1}z^{n+i-1}\right)_x = \sum_n np_n z^{1-n}\left(f_{,x} - \sum_{j=0}^{n-2} A_{j,x}z^j\right), \tag{4.9a}$$

$$\sum_{n,i} z^i(n+i)A_{n+i-1}p_{n,x} = \sum_n p_{n,x}z^{1-n}\sum_i A_{n+i-1}(n+i)z^{n+i-1}$$
$$= \sum_n p_{n,x}z^{1-n}\left[\sum_{i=0}^{\infty} A_i(i+1)z^i - \sum_{j=0}^{n-2} A_j(j+1)z^j\right]$$
$$= \sum_n p_{n,x}z^{1-n}\left[f + zf_{,z} - \sum_{j=0}^{n-2} A_j(j+1)z^j\right]; \tag{4.9b}$$

$$\sum_{i,n} z^i \frac{1}{2}\mathscr{D}_r(p_n)\mathscr{D}_r(A_{n+i-1}) = \frac{1}{2}\sum_n \mathscr{D}_r(p_n)z^{1-n}\sum_i \mathscr{D}_r(A_{n+i-1})z^{n+i-1}$$
$$= \frac{1}{2}\sum_n \mathscr{D}_r(p_n)z^{1-n}\mathscr{D}_r\left(f - \sum_{j=0}^{n-2} A_j z^j\right)$$
$$= -\frac{1}{2}\mathscr{D}_r(f)\sum_n \mathscr{D}_r(p_n)z^{1-n} \tag{4.9c}$$
$$-\frac{1}{2}\sum_n\sum_{j=0}^{n-2}\mathscr{D}_r(p_n)\mathscr{D}_r(A_j)z^{j+1-n}. \tag{4.9c'}$$

The expression (4.9c′) can be simplified:

Lemma 4.10:

$$\sum_{j<n-2} \mathscr{D}_r(p_n)\mathscr{D}_r(A_j)z^{j+1-n} = 0. \tag{4.11}$$

Proof: Denoting temporarily

$$\left(\sum_l \pi_l z^l\right)_- := \sum_{l<0} \pi_l z^l, \quad \pi_l \in \mathscr{A} \text{ for all } l, \tag{4.12}$$

and noticing that

$$\mathscr{D}_r(p_n) = \sum_i p_{n,i}\mathscr{D}_r(A_i), \tag{4.13}$$

we get

$$\sum_{j<n-2} \mathscr{D}_r(p_n)\mathscr{D}_r(A_j)z^{j+1-n}$$
$$= \Big(\sum_{j,n} \mathscr{D}_r(p_n)\mathscr{D}_r(A_j)z^{j+1-n}\Big)_-$$
$$= \Big[\sum_{i,j} \mathscr{D}_r(A_i)\mathscr{D}_r(A_j)\sum_n p_{n,i}z^{j+1-n}\Big]_- . \tag{4.14}$$

We will show that the expression $\Sigma_n\, p_{n,i}z^{j+1-n}$ is symmetric in (i,j). Since

$$\mathscr{D}_r(A_i)\mathscr{D}_r(A_j) = -\mathscr{D}_r(A_j)\mathscr{D}_r(A_i),$$

the expression inside the curly bracket in (4.14) will then vanish. Now,

$$\Big[\frac{\partial^k}{\partial A_{i_1}\cdots\partial A_{i_k}}\Big(\sum_n p_{n,i}z^{j+1-n}\Big)\Big]\Big|_{A=0} \quad \text{[by (4.3)]}$$
$$= \sum_n z^{j+1-n}\frac{\partial^{k+2}\overline{H}_m}{\partial A_{i_1}\cdots\partial A_{i_k}\partial A_i\partial A_n}\Big|_{A=0} \quad \text{[by (3.7)]}$$
$$= \sum_n z^{j+1-n}\frac{m!}{(m-k-1)!}$$
$$\times\delta\Big(m+2, \sum(i_s+2)+i+2+n+2\Big)$$
$$= \frac{m!}{(m-k-1)!}z^{j+1+\Sigma(i_s+2)-m+i+2},$$

which is indeed symmetric in (i,j). It remains to consider the case when $p_{n,i}$ is a constant. By formulas (4.3), (3.7), it happens when $n+i=m$, in which case $p_{n,i}=m$, so that

$$p_{n,i}z^{j+1-n} = mz^{j+1+i-m}$$

is again symmetric in (i,j). ■

Collecting together the expressions (4.9a)–(4.9c), we finally get

$$f_{,t} = f_{,x}\sum_n np_nz^{1-n} + (f+zf_{,z})\sum_n p_{n,x}z^{1-n}$$
$$-\frac{1}{2}\mathscr{D}_r(f)\sum_n \mathscr{D}_r(p_n)z^{1-n} \tag{4.15a}$$
$$-\sum_{j<n-2} z^{j+1-n}\,[np_nA_{j,x} + p_{n,x}A_j(j+1)]. \tag{4.15b}$$

Before proceeding further, we show that the expression (4.15b) is trivial:

Lemma 4.16: Denote

$$S = \sum_{j<n-2} z^{j+1-n}[np_nA_{j,x} + p_{n,x}A_j(j+1)]. \tag{4.17}$$

Then

$$S\sim 0. \tag{4.18}$$

Proof: We have,

$$S\sim \sum_{j<n-2} z^{j+1-n}(n-j-1)\,p_nA_{j,x}$$
$$= \sum_{j=0}^{m-2} A_{j,x}\sum_{n=j+2}^{m} z^{j+1-n}(n-j-1)\,p_n. \tag{4.19}$$

Hence, $S\sim 0$ iff

$$\frac{\partial}{\partial A_i}\Big[\sum_{n=j+2}^{m} z^{j+1-n}(n-j-1)\,p_n\Big]$$
$$= \frac{\partial}{\partial A_j}\Big[\sum_{n=i+2}^{m} z^{i+1-n}(n-i-1)\,p_n\Big], \tag{4.20}$$

for all $i,j\leqslant m-2$. This can be rewritten with the help of the notation (4.12) as

$$\Big[\sum_n z^{j+1-n}(n-j-1)\,p_{n,i}\Big]_-$$
$$= \Big[\sum_n z^{i+1-n}(n-i-1)\,p_{n,j}\Big]_- . \tag{4.21}$$

To show that formula (4.21) is true, we first apply the operator $(\partial^k/(\partial A_{i_1}\cdots\partial A_{i_k}))|_{A=0}$ to the lhs of (4.21) and verify that the result is symmetric in (i,j). By (4.6), we get

$$\Big\{\Big[\sum_n z^{j+1-n}(n-j-1)\frac{\partial^k}{\partial A_{i_1}\cdots\partial A_{i_k}}\frac{\partial}{\partial A_i}\frac{\partial \overline{H}_m}{\partial A_n}\Big]\Big|_{A=0}\Big\}_- \quad \text{[by (3.7)]}$$
$$= \Big\{\sum_n z^{j+1-n}(n-j-1)\delta\Big(m+2, \sum(i_s+2)+i+2+n+2\Big)\frac{m!}{(m-k-1)!}\Big\}_-$$
$$= \Big\{\frac{m!}{(m-k-1)!}z^{j+1+\Sigma(i_s+2)+2-m+i}\Big[m-2-\sum(i_s+2)-i-j-1\Big]\Big\}_-,$$

and this expression is indeed symmetric in (i,j). Finally, the constant terms in the lhs of (4.21) occur when $n+i=m$, in which case $p_{m-i,i}=m$ by formulas (4.6) and (3.7), and the constant term becomes

$$\{z^{j+1-m+i}(j+1-m+i)m\}_-,$$

which is again symmetric in (i,j). ■

We now are ready to construct the desired infinity of conserved densities of the equation (4.15).

Theorem 4.22: Let the sequence of operators

$$E_i \in \mathbb{Q}\Big[z, \frac{d}{dz}\Big], \quad i\in\mathbf{Z}_+, \tag{4.23}$$

be given by the formulas

$$E_i = E_0(z^2\partial_z + (N/2)z)^i z^{i+1}, \tag{4.24}$$

where E_0 is arbitrary. Then

$$\left[\sum_{i=0}^{\infty} E_i\left[\frac{f^{i+1}}{(i+1)!}\right]\right]_{,t}$$
$$= \left[\sum_n np_n \sum_{i=0}^{\infty} E_i\left(\frac{f^{i+1}}{(i+1)!}z^{1-n}\right)\right]_{,x}$$
$$-\tfrac{1}{2}\mathscr{D}_r\left[\sum_n \mathscr{D}_r(p_n)\sum_{i=0}^{\infty} E_i\left(\frac{f^{i+1}}{(i+1)!}\right.\right.$$
$$\left.\left.\times z^{1-n}\right)\right] - E_0(S). \tag{4.25}$$

Remark 4.26: Since E_0 commutes with ∂, by formula (4.18) $E_0(S)\sim 0$. Thus, formula (4.25) implies that all the coefficients in the generating series

$$\sum_{i=0}^{\infty} E_i\left(\frac{f^{i+1}}{(i+1)!}\right), \tag{4.27}$$

are conserved densities for the system (4.5). (Obviously, all the series in formula (4.25) converge z-adically.)

Proof of Theorem 4.22: We have,

$$\left[\sum_i E_i\left(\frac{f^{i+1}}{(i+1)!}\right)\right]_{,t} = \sum_i E_i\left(\frac{f^i}{i!}f_{,t}\right) \text{ [by (4.15)]}$$
$$= \sum_i E_i\left\{\frac{f^i}{i!}\left[f_{,x}\sum_n np_n z^{1-n} + (f+zf_{,z})\sum_n p_{n,x}z^{1-n} - \tfrac{1}{2}\mathscr{D}_r(f)\sum_n \mathscr{D}_r(p_n)z^{1-n}\right]\right\} - \tag{4.28a}$$
$$-\sum_i E_i\left\{\frac{f^i}{i!}\sum_{j<n-2} z^{j+1-n}[np_n A_{j,x} + p_{n,x}A_j(j+1)]\right\}. \tag{4.28b}$$

First, we transform each of the three summands in the sum (4.28a). We have,

$$\sum_i E_i\left(\frac{f^i}{i!}f_{,x}\sum_n np_n z^{1-n}\right)$$
$$= \left[\sum_n np_n\sum_i E_i\left(\frac{f^{i+1}}{(i+1)!}z^{1-n}\right)\right]_{,x} \tag{4.29}$$
$$-\sum_n np_{n,x}\sum_i E_i\left(\frac{f^{i+1}}{(i+1)!}z^{1-n}\right); \tag{4.30}$$

$$\sum_i E_i\left[\frac{f^i}{i!}(f+zf_{,z})\sum_n p_{n,x}z^{1-n}\right]$$
$$= \sum_n p_{n,x}\sum_i E_i\left[\frac{f^{i+1}}{i!}z^{1-n} + z^{1-n}z\partial_z\left(\frac{f^{i+1}}{(i+1)!}\right)\right]; \tag{4.31}$$

$$-\tfrac{1}{2}\sum_i E_i\left[\frac{f^i}{i!}\mathscr{D}_r(f)\mathscr{D}_r(p_n)z^{1-n}\right]$$
$$= -\tfrac{1}{2}\mathscr{D}_r\left\{\sum_i E_i\left[\frac{f^{i+1}}{(i+1)!}\sum_n \mathscr{D}_r(p_n)z^{1-n}\right]\right\} \tag{4.32}$$
$$+\frac{N}{2}\sum_n p_{n,x}\sum_i E_i\left(\frac{f^{i+1}}{(i+1)!}z^{1-n}\right). \tag{4.33}$$

Second, the expression (4.28b) can be transformed as

$$-E_0(S) \tag{4.34}$$
$$-\sum_{j<n-2}[np_nA_{j,x} + p_{n,x}A_j(j+1)]\sum_i E_{i+1}$$
$$\times\left(\frac{f^{i+1}}{(i+1)!}z^{j+1-n}\right). \tag{4.35}$$

The expressions (4.29), (4.32), and (4.34) combine into the desired formula (4.25). We are going to show that the remaining expressions, (4.30), (4.31), (4.33), and (4.35), add up to zero. Now, each of these expressions is linear in $A_{q,x}$, $0<q<m-2$. Hence, for each q, picking out the ceofficients in front of $A_{q,x}$ and then considering this coefficient as an operator acting on $f^{i+1}/(i+1)!$ for each fixed i, it will be enough to check the following operator identity:

$$-\sum_n np_{n,q}E_iz^{1-n} + \sum_n p_{n,q}E_iz^{1-n}[(i+1)+z\partial_z]$$
$$+\frac{N}{2}\sum_n p_{n,q}E_iz^{1-n} \tag{4.36a}$$
$$-\sum_{n>q+2} np_nE_{i+1}z^{q+1-n}$$
$$-\sum_{j<n-2} p_{n,q}A_j(j+1)E_{i+1}z^{j+1-n} = 0. \tag{4.36b}$$

Lemma 4.37:

$$E_{i+1} = E_i(N/2+i+z\partial_z)z^2. \tag{4.38}$$

Granted the Lemma, which we shall prove later on, the expression (4.36b) can be transformed into

$$-\sum_{n>q+2} np_nE_i(N/2+i+z\partial_z)z^{q+3-n}$$
$$-\sum_{j<n-2}\{p_{n,q}A_j(j+1)E_i\left(\frac{N}{2}+i+z\partial_z\right)z^{j+3-n}. \tag{4.39}$$

The last transformation we perform on the second of the three summands in the expression (4.36a):

$$\sum_n p_{n,q}E_iz^{1-n}[(i+1)+z\partial_z]$$
$$= \sum_n p_{n,q}E_i[(n+i)-z\partial_z]z^{1-n}. \tag{4.40}$$

Now, collecting together terms of the form $E_iz\partial_z$ from

(4.39) and (4.40), we arrive at the following identity to be verified:

$$\sum_n p_{n,q} z^{1-n} = \sum_{n>q+2} n p_n z^{q+3-n} + \sum_{j<n-2} \{p_{n,q} A_j (j+1) z^{j+3-n}\}; \quad (4.41)$$

terms proportional to $(N/2)E_i$ in (4.36a) and (4.39) lead to the same identity (4.41); finally, after adding the first summand in (4.36a) to (4.40), the remaining terms, all proportional to $E_i i$, amount again to the same identity (4.41).

Denoting

$$\Big(\sum_\ell \pi_\ell z^\ell\Big)_{<\gamma} := \sum_{\ell<\gamma} \pi_\ell z^\ell, \quad \pi_\ell \in \mathscr{A} \text{ for all } \ell, \quad (4.42)$$

the identity (4.41) can be rewritten in the form

$$\Big(\sum_n p_{n,q} z^{1-n}\Big)_{<1} = \Big[\sum_n n p_n z^{q+3-n} + \sum_{n,j} \{p_{n,q} A_j (j+1) \times z^{j+3-n}\}\Big]_{<1}. \quad (4.43)$$

To prove formula (4.43), we first apply the operator $((\partial^k/\partial A_{i_1} \cdots \partial A_{i_k}))|_{A=0}$ to each of the three terms in it. Using formulas (4.6) and (3.7) we get:

$$\begin{aligned}
&\Big\{\Big[\sum_n z^{1-n} \frac{\partial^k}{\partial A_{i_1}\cdots\partial A_{i_k}} \frac{\partial}{\partial A_q}\Big(\frac{\partial \bar{H}_m}{\partial A_n}\Big)\Big]\Big|_{A=0}\Big\}_{<1} \\
&= \Big[\sum_n z^{1-n} \frac{m!}{(m-k-1)!} \delta\Big(m+2, \sum (i_s+2) + q + 2 + n + 2\Big)\Big]_{<1} \\
&= \frac{m!}{(m-k-1)!} (z^{1-m+\Sigma(i_s+2)+q+2})_{<1}, \qquad (4.44\text{L})
\end{aligned}$$

$$\begin{aligned}
&\Big\{\Big[\sum_n n z^{q+3-n} \frac{\partial^k}{\partial A_{i_1}\cdots\partial A_{i_k}} \Big(\frac{\partial \bar{H}_m}{\partial A_n}\Big)\Big]\Big|_{A=0}\Big\}_{<1} \\
&= \Big[\sum_n n z^{q+3-n} \frac{m!}{(m-k)!} \delta\Big(m+2, \sum (i_s+2) + n + 2\Big)\Big]_{<1} \\
&= \frac{m!}{(m-k)!} \Big[m - \sum (i_s+2)\Big] (z^{q+3-m+\Sigma(i_s+2)})_{<1}, \qquad (4.44\text{Ra})
\end{aligned}$$

$$\begin{aligned}
&\Big\{\Big[\sum_{n,j} (j+1) z^{j+3-n} \frac{\partial^k}{\partial A_{i_1}\cdots\partial A_{i_k}} \Big(A_j \frac{\partial^2 \bar{H}_m}{\partial A_n \partial A_q}\Big)\Big]\Big|_{A=0}\Big\}_{<1} \\
&= \Big\{\Big[\sum_{\xi=1}^k (i_\xi+1) \sum_n z^{i_\xi+3-n} \frac{\partial^{k-1}}{\partial A_{i_1}\cdots\hat{\partial} A_{i_\xi}\cdots\partial A_{i_k}} \frac{\partial}{\partial A_n}\Big(\frac{\partial \bar{H}_m}{\partial A_q}\Big)\Big]_{A=0}\Big\}_{<1} \\
&= \Big[\sum_\xi (i_\xi+1) \sum_n z^{i_\xi+3-n} \frac{m!}{(m-k)!} \delta\Big(m+2, \sum (i_s+2) - (i_\xi+2) + n + 2 + q + 2\Big)\Big]_{<1} \\
&= \frac{m!}{(m-k)!} \sum_\xi (i_\xi+1) (z^{3+\Sigma(i_s+2)-m+q})_{<1} \\
&= \frac{m!}{(m-k)!} \Big[\sum (i_s+2) - k\Big] (z^{3+\Sigma(i_s+2)-m+q})_{<1}. \qquad (4.44\text{Rb})
\end{aligned}$$

Adding up the expressions (4.44Ra) and (4.44Rb) we obtain the expression (4.44L).

Second, to compare the constant terms in formula (4.43), we apply the operation "$|_{A=0}$" to it. From the lhs we obtain

$$\Big(\sum_n \frac{\partial^2 \bar{H}_m}{\partial A_q \partial A_n}\Big|_{A=0} z^{1-n}\Big)_{<1} = m(z^{1-(m-q-2)})_{<1}, \quad (4.45\text{L})$$

and from the rhs we get

$$\Big(\sum_n n \frac{\partial \bar{H}_m}{\partial A_n}\Big|_{A=0} z^{q+3-n}\Big)_{<1} = m(z^{q+3-m})_{<1}, \quad (4.45\text{R})$$

which is the same as (4.45L). Theorem 4.22 proved, modulo Lemma 4.37. ■

Proof of Lemma 4.37: Take the obvious identity

$$z^2\partial_z + (N/2)z = z^{i+1}(N/2 + i + z\partial_z)z^{-i}, \quad (4.46\text{b})$$

and multiply it by $E_0(z^2\partial_z + (N/2)z)^i$ from the left, and by z^{i+2} from the right, resulting in

$$E_0(z^2\partial_z + (N/2)z)^{i+1} = E_0(z^2\partial_z + (N/2)z)^i z^{i+1} \times ((N/2) + i + z\partial_z))z^2,$$

which is the desired formula (4.38) in view of the definition (4.24). ■

Thus, we have two constructions for the infinity of conserved densities: the explicit formula (3.22) which has been proven only for the flow #2, and the generating series formula (4.27) which applies to every flow. We now establish the connection between these two constructions.

Theorem 4.47:

$$\sum_{i=0}^{\infty} E_i\left(\frac{f^{i+1}}{(i+1)!}\right)=E_0\left(\sum_{i=0}^{\infty} H_i z^{i+1}\right). \tag{4.48}$$

Proof: Since E_0 is arbitrary, we can use formula (4.24) to transform formula (4.48) into the equivalent form:

$$\sum_{i=0}^{\infty}\left(z^2\partial_z+\frac{N}{2}z\right)^i\left(\frac{z^{i+1}f^{i+1}}{(i+1)!}\right)=\sum_{i=0}^{\infty} H_i z^{i+1}. \tag{4.49}$$

Denote

$$E:=z^2\partial_z+(N/2)z, \tag{4.50}$$

$$\phi:=zf=\sum_{i=0}^{\infty} A_i z^{i+1}. \tag{4.51}$$

Then the identity (4.49) takes the form

$$\sum_i E^i\left(\frac{\phi^{i+1}}{(i+1)!}\right)=\sum_i H_i z^{i+1}. \tag{4.52}$$

To verify formula (4.52), we check that both of its sides yield the same results when we apply the operations: first, $|_{A=0}$; second, $(\partial/\partial A_j)|_{A=0}$; and third, $((\partial^k/\partial A_{j_1}\cdots\partial A_{j_k}))|_{A=0}$, $k\geqslant 2$. We have: (1) When $A=0$, both ϕ and H vanish, so we get $0=0$.

(2) Further,

$$\frac{\partial}{\partial A_j}\left[\sum_i E^i\left(\frac{\phi^{i+1}}{(i+1)!}\right)\right]=\sum_i E^i\left(\frac{\phi^i}{i!}z^{j+1}\right), \tag{4.53}$$

so the lhs becomes z^{j+1}, which is the same as

$$\sum_i \frac{\partial H_i}{\partial A_j}\bigg|_{A=0} z^{i+1}=\sum_i c_{i|1}\delta(i+2,j+2)z^{i+1}=z^{j+1}.$$

(3) Now, for $k\geqslant 2$,

$$\left\{\frac{\partial^k}{\partial A_{j_1}\cdots\partial A_{j_k}}\left[\sum_i E^i\left(\frac{\phi^{i+1}}{(i+1)!}\right)\right]\right\}\bigg|_{A=0} \quad [\text{by } (4.53)]$$

$$=\left[\sum_i E^i\left(\frac{\phi^{i+1-k}}{(i+1-k)!}\prod_s z^{j_s+1}\right)\right]\bigg|_{A=0}$$

$$=E^{k-1}(z^{\Sigma(j_s+1)}). \tag{4.54}$$

But

$$E(z^\rho)\ [\text{by } (4.50)]=(\rho+N/2)z^{\rho+1},$$

so that

$$E^\eta(z^\rho)=\prod_{s=1}^{\eta}(\rho+N/2+s-1)z^{\rho+\eta},\quad \forall\eta\in\mathbb{N}. \tag{4.55}$$

Hence, denoting

$$\rho:=\sum_{s=1}^{k}(j_s+1)=\sum_{s=1}^{k}(j_s+2)-k, \tag{4.56}$$

the expression (4.54) becomes

$$z^{\rho+k-1}\prod_{s=1}^{k-1}\left(\rho+\frac{N}{2}+s-1\right). \tag{4.57}$$

On the other hand,

$$\left[\frac{\partial^k}{\partial A_{j_1}\cdots\partial A_{j_k}}\left(\sum_i H_i z^{i+1}\right)\right]\bigg|_{A=0} \quad [\text{by } (3.22)]$$

$$=\sum_i z^{i+1}c_{i|k}\delta\left(i+2,\sum(j_s+2)\right)$$

$$=z^{i+1}c_{i|k}|_{i=\Sigma(j_s+2)-2}\quad [\text{by } (4.56)]$$

$$=z^{\rho+k-2+1}c_{\rho+k-2|k}. \tag{4.58}$$

Denoting

$$m:=\rho+k-2,$$

and comparing expressions (4.57) and (4.58), we see that they are equal provided the identity

$$c_{m|k}=\prod_{s=1}^{k-1}\left(m-k+2+\frac{N}{2}+s-1\right),\quad k\geqslant 2, \tag{4.59}$$

holds. To show that it does, we transform the product as:

$$\prod_{s=1}^{k-1}\left[m-(k-1-s)+\frac{N}{2}\right]$$

$$=\prod_{s=0}^{k-2}\left(m-s+\frac{N}{2}\right)$$

$$=\begin{cases} m+N/2, & k=2\\ \left[\prod_{s=0}^{k-3}\left(m-s+\frac{N}{2}\right)\right]\left(m-k+2+\frac{N}{2}\right), & k>2,\end{cases}$$

and this is equivalent to formula (3.23).

ACKNOWLEDGMENT

Discussions with John Harnad were helpful to me. I thank the Institute for Advanced Study for its hospitality. This work was partially supported by the National Science Foundation.

[1]D. J. Benney, Stud. Appl. Math. **52**, 45–50 (1973).
[2]B. A. Kupershmidt and Yu. I. Manin, Funct. Anal. Appl. **11** (3), 31–42 (1977) (Russian); 188–197 (English).
[3]B. A. Kupershmidt and Yu. I. Manin, Funct. Anal. Appl. **12** (1), 25–37 (1987) (Russian); 20–29 (English).
[4]D. R. Lebedev and Yu. I. Manin, Phys. Lett. A **74**, 154–156 (1979).
[5]Yu. I. Manin, J. Sov. Math. **11**, 5–152 (1978) (Russian); 1–122 (English).
[6]B. A. Kupershmidt, Comm. Math. Phys. **99**, 51 (1985).
[7]G. Wilson, Math. Proc. Cambridge Philos. Soc. **86**, 131 (1979).
[8]B. A. Kupershmidt, *Elements of Superintegrable Systems* (Reidel, Dordrecht, Boston, 1987).
[9]B. A. Kupershmidt, Proc. R. Irish Acad. A **83**, 45 (1983).
[10]Yu. I. Manin and A. O. Radul, Comm. Math. Phys. **98**, 65 (1985).

Hamiltonian structure of the super evolution equation

Yi-shen Li
Department of Mathematics, University of Science and Technology of China, Hefei, Anhui, People's Republic of China
Li-ning Zhang
Institute of Theoretical Physics, Academia Sinica, Beijing, People's Republic of China

(Received 28 October 1988; accepted for publication 16 August 1989)

The constrained variational calculus proposed in previous papers [Y. Zheng, Y. Li, and D. Chen, Sci. Sinica A **24**, 138 (1986); G. Tu, Kexue Tangbao **29**, 1227 (1984)] is generalized to the supersymmetric case. Utilizing this method to some super AKNS system, which has soliton solutions and conserved quantities, their equations of motion with the Hamiltonian structure can be written as a 4×4 matrix. This is the supersymmetric symplectic matrix and h_n are Hamiltonians of the system. The conserved quantities worked out in recent literature are just the first few terms of our analytic expression.

I. INTRODUCTION

Soliton theory has achieved great success during the last two decades, and its recent development is still quite exciting. Now, the supersymmetric extension of this theory has become more and more fashionable and has attracted much attention among mathematicians as well as theoretical physicists.[1–3] The algebraic structure of the soliton equations and their supersymmetric generalization, undoubtedly, plays a special important role.

Recently, Choudhury and Roy[4] considered the Hamiltonian structure as well as Bäcklund transformation of the superevolution equation for a super AKNS system, but it is much regretted that their work does not contain a complete Hamiltonian formulation of the super AKNS scheme. Furthermore, the consistency between their assumption $\beta=0$ and the three-component equations has to be checked. In this paper, we consider the super AKNS scheme and give the Hamiltonian formulation completely, by using the method that we called constrained variational calculus in our previous papers.[5,6]

We believe our method is also valid for other integrable supersymmetric systems as well. In addition, we have studied the Lie–Bäcklund symmetries, as well as other symmetries with fruitful structures and their algebraic property. We shall publish these results in a forthcoming paper.

II. FORMULATION OF THE SUPER AKNS SCHEME

In our previous paper,[7] we have considered the eigenvalue problem

$$q_x=Mq,\quad q=\begin{pmatrix} q_1\\ q_2\\ q_3\end{pmatrix},\quad M=\begin{pmatrix} -\xi & q & \alpha\\ r & \xi & \beta\\ -\beta & \alpha & 0\end{pmatrix},\tag{2.1}$$

where ξ, q, r are even: $P(\xi)=p(q)=p(r)=0$; α,β are odd: $p(\alpha)=p(\beta)=0$, and ξ is a constant eigenparameter; q, r, α, β are functions of x,t.

Associated with (2.1) is a temporal evolution equation

$$q_t=Nq,\quad N=\begin{pmatrix} A & B & \rho\\ C & -A & \delta\\ -\delta & \rho & 0\end{pmatrix},\tag{2.2}$$

where $p(A)=p(B)=p(C)=0$, $p(\rho)=p(\delta)=1$. From the compatibility condition $\varphi_{xt}=\varphi_{tx}$, we obtain

$$M_t-N_x+MN-NM=0.\tag{2.3}$$

We can write (2.3) as follows:

$$\begin{aligned} &-A_x+qC-rB-\alpha\delta-\beta\rho=0,\\ &q_t-B_x-2B\xi-2qA+2\alpha\rho=0,\\ &r_t-C_x+2C\xi+2rA-2\beta\delta=0,\\ &\alpha_t-\rho_x-\rho\xi-\alpha A+q\delta-\beta B=0,\\ &\beta_t-\delta_x+\delta\xi+\beta A+\gamma\rho-\alpha C=0.\end{aligned}\tag{2.4}$$

We put A, B, C, ρ, δ to be polynomial of ξ of order n

$$\begin{aligned} &A=\sum_{j=0}^{n}a_j\xi^{n-j},\quad B=\sum_{j=1}^{n}b_j\xi^{n-j},\quad C=\sum_{j=1}^{n}c_j\xi^{n-j},\\ &\rho=\sum_{j=1}^{n}\rho_j\xi^{n-j},\quad \delta=\sum_{j=1}^{n}\delta_j\xi^{n-j}.\end{aligned}\tag{2.5}$$

Substituting Eq. (2.5) into Eq. (2.4) and equating the coefficients of ξ^n to be zero, we get equations for a_j, b_j, c_j, α_j, β_j. Solving them with $a_0=-1$, and $a_j=0(j\geqslant 1)$ we obtain a super AKNS hierarchy. We denote the special solution (A,v) as $(\overline{A},\overline{v})$. Then we have

$$U_t=2L^nU_0=2\overline{V}_{n+1},\quad n=0,1,2,\ldots,\tag{2.6}$$

where

 0022-2488/90/020470-06$03.00

$$U=\begin{pmatrix} r \\ -q \\ 2\beta \\ -2\alpha \end{pmatrix}, \quad U_0=\begin{pmatrix} r \\ q \\ \beta \\ \alpha \end{pmatrix}^{1/2}$$

$$L=\frac{1}{2}\begin{pmatrix} D-2rD^{-1}q & 2rD^{-1}r & 2rD^{-1}\alpha+2\beta & 2rD^{-1}\beta \\ -2qD^{-1}q & -D+2qD^{-1}r & 2qD^{-1}\alpha & 2qD^{-1}\beta+2\alpha \\ -2\beta D^{-1}q+2\alpha & 2\beta D^{-1}q & 2D+2\beta D^{-1}\alpha & 2\beta D^{-1}\beta-2r \\ -2\alpha D^{-1}q & 2\alpha D^{-1}r-2\beta & 2\alpha D^{-1}\alpha+2q & -2D+2\alpha D^{-1}\beta \end{pmatrix}. \tag{2.7}$$

When $n=0,1,2$, Eq. (2.6) reads:

$n=0$ case

$$\begin{pmatrix} r \\ q \\ \beta \\ \alpha \end{pmatrix}_t=\begin{pmatrix} 2r \\ -2q \\ \beta \\ -\alpha \end{pmatrix}, \tag{2.8}$$

since

$$2U_0=2\overline{V}_1=2\begin{pmatrix} r \\ q \\ \beta \\ \alpha \end{pmatrix}, \tag{2.9}$$

$n=1$ case

$$\begin{pmatrix} r \\ q \\ \beta \\ \alpha \end{pmatrix}_t=\begin{pmatrix} r_x \\ q_x \\ \beta_x \\ \alpha_x \end{pmatrix}, \tag{2.10}$$

since

$$2LU_0=2\overline{V}_2=2\begin{pmatrix} \frac{1}{2}r_x \\ -\frac{1}{2}q_x \\ \beta_x \\ -\alpha_x \end{pmatrix}; \tag{2.11}$$

and $n=2$ case

$$\begin{pmatrix} r \\ q \\ \beta \\ \alpha \end{pmatrix}_t=\begin{pmatrix} \frac{1}{2}r_{xx}-qr^2+2r\alpha\beta+2\beta\beta_x \\ -\frac{1}{2}q_{xx}+q^2r-2q\alpha\beta+2\alpha\alpha_x \\ \beta_{xx}-\frac{1}{2}\beta qr+r\alpha_x+\frac{1}{2}\alpha r_x \\ -\alpha_{xx}+\frac{1}{2}\alpha qr-q\beta_x-\frac{1}{2}\beta q_x \end{pmatrix}, \tag{2.12}$$

since

$$2L^2U_0=2V_3=2\begin{pmatrix} \frac{1}{4}r_{xx}-\frac{1}{2}qr^2+r\alpha\beta+\beta\beta_x \\ \frac{1}{4}q_{xx}-\frac{1}{2}qr^2+q\alpha\beta-\alpha\alpha_x \\ \beta_{xx}-\frac{1}{2}\beta qr+r\alpha_x+\frac{1}{2}\alpha r_x \\ \alpha_{xx}-\frac{1}{2}\alpha qr+q\beta_x+\frac{1}{2}\beta q_x \end{pmatrix}. \tag{2.13}$$

III. FUNDAMENTAL EQUATION

In this section, we follow the notations of Ref. 7 and assume that $d_0=+1$, $d_j=0$, $j=1,2,\ldots,n$, then Eqs. (2.6), (2.9), and (2.13) read

$$U_t=2L^jU_0, \tag{3.1}$$

and

$$V_1=U_0, \quad V_j=L^{j-1}U_0, \tag{3.2}$$

respectively.

The special solution (A,v) has an alternative expression in Ref. 4; we have proved that

$$\overline{A}=-1+\bar{a}_1/\xi+\bar{a}_2/\xi^2+\cdots, \tag{3.3a}$$

$$\begin{pmatrix} C \\ B \\ \delta \\ \rho \end{pmatrix}\equiv\overline{V}=\bar{v}_1/\xi+\bar{v}_2/\xi^2+\cdots, \quad \overline{V}_j=\begin{pmatrix} \bar{c}_j \\ \bar{b}_j \\ \bar{\delta}_j \\ \bar{\rho}_j \end{pmatrix}. \tag{3.3b}$$

Then $\overline{A}$ and $\overline{V}$ will satisfy the following equations:

$$\begin{aligned} &-\overline{A}_x+q\overline{C}-r\overline{B}-\alpha\bar{\delta}-\beta\bar{\rho}=0,\\ &-\overline{B}_x-2\overline{B}\xi-2q\overline{A}+2\alpha\bar{\rho}=0,\\ &-\overline{C}_x+2\overline{C}\xi+2r\overline{A}-2\beta\bar{\delta}=0,\\ &-\bar{\rho}_x-\bar{\rho}\xi+q\bar{\delta}-\alpha\overline{A}-\beta\overline{B}=0,\\ &-\bar{\delta}_x+\bar{\delta}\xi+r\bar{\rho}+\beta\overline{A}-\alpha\overline{C}=0. \end{aligned} \tag{3.4}$$

From Eq. (3.4) we obtain

$$-(\overline{A}^2+\overline{BC})_x+2(\bar{\rho}\bar{\delta})_x=0.$$

Integrating with respect to x, we have

$$-\overline{A}^2-\overline{BC}+2\bar{\rho}\bar{\delta}=1. \tag{3.5}$$

Substituting Eq. (3.3) into Eq. (3.5), we obtain

$$\bar{a}_1=0, \bar{a}_j=\frac{1}{2}\sum_{l=1}^{j-1}\bar{b}_l\overline{C}_{j-l}+\frac{1}{2}\sum_{l=1}^{j-1}\bar{a}_l\bar{a}_{j-l}-\sum_{l=1}^{j-1}\bar{\rho}_l\bar{\delta}_{j-1}. \tag{3.6}$$

IV. FUNDAMENTAL EQUATION—AN ALTERNATIVE FORM

For the sake of convenience to compare with the Hamiltonian form, we need to reformulate Eq. (3.7) in an alternative form, where we use a new matrix

$$L_1=\frac{1}{2}\begin{pmatrix} -D+2qD^{-1}r & -2qD^{-1}q & qD^{-1}\beta+\alpha & -qD^{-1}\alpha \\ 2rD^{-1}r & D-2rD^{-1}q & rD^{-1}\beta & -rD^{-1}\alpha-\beta \\ 4\alpha D^{-1}r-4\beta & -4\alpha D^{-1}q & -2D+2\alpha D^{-1}\beta & -2\alpha D^{-1}\alpha-2q \\ -4\beta D^{-1}r & 4\beta D^{-1}q-4\alpha & -2\beta D^{-1}\beta+2r & 2D+2\beta D^{-1}\alpha \end{pmatrix}, \tag{4.1}$$

as well as

$$\tilde{U}=\begin{pmatrix} r\\ q\\ \beta\\ \alpha\end{pmatrix} = -\frac{1}{2}gv=\frac{1}{2}\begin{pmatrix} 2&0&0&0\\ 0&-2&0&0\\ 2&0&1&0\\ 0&0&0&-1\end{pmatrix}\begin{pmatrix} r\\ -q\\ 2\beta\\ -2\alpha\end{pmatrix}, \tag{4.2}$$

$$\tilde{U}_0=\begin{pmatrix} -\frac{1}{2}q\\ -\frac{1}{2}r\\ -\frac{1}{2}\alpha\\ \beta\end{pmatrix}=J^{-1}gU=J^{-1}g\begin{pmatrix} r\\ q\\ \beta\\ \alpha\end{pmatrix}, \tag{4.3}$$

$$J=\begin{pmatrix} 0&-4&0&0\\ 4&0&0&0\\ 0&0&0&1\\ 0&0&1&0\end{pmatrix},$$

$$g=\begin{pmatrix} 2&0&0&0\\ 0&-2&0&0\\ 0&0&1&0\\ 0&0&0&-1\end{pmatrix}, \tag{4.4}$$

$$J^{-}=\begin{pmatrix} 0&-\frac{1}{4}&0&0\\ -\frac{1}{4}&0&0&0\\ 0&0&0&1\\ 0&0&1&0\end{pmatrix}.$$

It is easy to get

$$L_1=J^{-1}gLg^{-1}L,$$

and hence

$$L_1^n=J^{-1}gL^n g^{-1}L.$$

Then the equation (2.6), $U_t=2L^nU_0$, becomes

$$\widetilde{U}_t=JL_1^n\widetilde{U}_0=J\begin{pmatrix} -\frac{1}{2}b_{n+1}\\ -\frac{1}{2}c_{n+1}\\ -\rho_{n+1}\\ \delta_{n+1}\end{pmatrix}. \tag{4.5}$$

This is another form of the fundamental equation, where

$$L_1\widetilde{U}_0=\begin{pmatrix} \frac{1}{4}q_x\\ -\frac{1}{4}r_x\\ \alpha_x\\ \beta_x\end{pmatrix},$$

$$L_1^2\widetilde{U}_0=\begin{pmatrix} -\frac{1}{8}q_{xx}+\frac{1}{4}q^2r-\frac{1}{2}q\alpha\beta+\frac{1}{2}\alpha\alpha_x\\ -\frac{1}{8}r_{xx}+\frac{1}{4}qr^2-\frac{1}{2}r\alpha\beta-\frac{1}{2}\beta\beta_x\\ -\alpha_{xx}+\frac{1}{2}qr\alpha-q\beta_x-\frac{1}{2}q_x\beta\\ \beta_{xx}-\frac{1}{2}qr\beta+r\alpha_x+\frac{1}{2}r_x\alpha\end{pmatrix}. \tag{4.6}$$

Therefore, the fundamental equation (3.7) can be rewritten as

$$\widetilde{U}_t\equiv\begin{pmatrix} r\\ q\\ \beta\\ \alpha\end{pmatrix}_t=JL_1^n\widetilde{U}_0=JL_1^n\begin{pmatrix} -\frac{1}{2}q\\ -\frac{1}{2}r\\ -\alpha\\ \beta\end{pmatrix}\equiv JL_1^n\widetilde{U}_0. \tag{4.7}$$

V. CONNECTION BETWEEN CONSTANTS OF MOTION AND EQUATION OF MOTION

In Ref. 7 it has been proved that the equation $U_t=2L^nU_0$, or $\widetilde{U}_t=JL_1^n\widetilde{U}_0$, have an infinite number of constants of motion as follows:

$$h_{n-1}=\int_{-\infty}^{\infty}(qf_n+\alpha g_n)dx,\quad n=1,2,3,..., \tag{5.1}$$

where

$$f_1=-\tfrac{1}{2}r,\ g_1=\beta,$$

$$f_{n+1}=\tfrac{1}{2}\Big(-r\delta_{n,0}+f_{n,x}-\beta g_n+g_n+q\sum_{l=1}^{n-1}f_lf_{n-l}+\alpha\sum_{l=1}^{n-1}f_lg_{n-l}\Big), \tag{5.2}$$

$$g_{n+1}=\Big[\beta\delta_{n,0}+g_{n,x}-\alpha f_n+q\sum_{l=1}^{n-1}f_lg_{n-l}\Big].$$

The first few h_n read

$$h_0=\int_{-\infty}^{\infty}\Big(-\frac{1}{2}qr+\alpha\beta\Big)dx,$$

$$h_1=\int_{-\infty}^{\infty}\Big(-\frac{1}{4}qr_x+\alpha\beta_x\Big)dx, \tag{5.3}$$

$$h_2=\int_{-\infty}^{\infty}\Big\{\frac{1}{8}(-qr_{xx}+q^2r^2-4q\beta\beta_x)+\alpha\beta_{xx}+\frac{1}{2}\alpha\alpha_xr-\frac{1}{2}qr\alpha\beta\Big\}dx.$$

It is easy to verify that

$$\begin{pmatrix} \frac{\delta}{\delta r}h_0\\ \frac{\delta}{\delta q}h_0\\ \frac{\delta}{\delta\beta}h_0\\ \frac{\delta}{\delta\alpha}h_0\end{pmatrix}=\begin{pmatrix} -\frac{1}{2}q\\ -\frac{1}{2}r\\ -\alpha\\ \beta\end{pmatrix}=\begin{pmatrix} -\frac{1}{2}b_1\\ -\frac{1}{2}c_1\\ -\rho_1\\ \delta_1\end{pmatrix}=\widetilde{U}_0, \tag{5.4}$$

$$\begin{pmatrix} \frac{\delta}{\delta r} & h_1 \\ \frac{\delta}{\delta q} & h_1 \\ \frac{\delta}{\delta \beta} & h_1 \\ \frac{\delta}{\delta \alpha} & h_1 \end{pmatrix} = \begin{pmatrix} \frac{1}{4}q \\ -\frac{1}{4}r_x \\ \alpha_x \\ \beta_x \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}b_2 \\ -\frac{1}{2}c_2 \\ -\rho_2 \\ \delta_2 \end{pmatrix} = L_1 \widetilde{U}_0, \qquad (5.5)$$

$$\begin{pmatrix} \frac{\delta}{\delta r} & h_2 \\ \frac{\delta}{\delta q} & h_2 \\ \frac{\delta}{\delta \beta} & h_2 \\ \frac{\delta}{\delta \alpha} & h_2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{8}q_{xx} + \frac{1}{4}q^2 r + \frac{1}{2}\alpha\alpha_x - \frac{1}{2}q\alpha\beta \\ -\frac{1}{8}r_{xx} + \frac{1}{4}qr^2 - \frac{1}{2}\beta\beta_x - \frac{1}{2}r\alpha\beta \\ -q\beta_x - \frac{1}{2}q_x\beta - \alpha_{xx} + \frac{1}{2}qr\alpha \\ \beta_{xx} + \alpha_x r + \frac{1}{2}\alpha r_x - \frac{1}{2}qr\beta \end{pmatrix}$$

$$= \begin{pmatrix} -\frac{1}{2}b_3 \\ -\frac{1}{2}c_3 \\ -\rho_3 \\ \delta_3 \end{pmatrix} = L_1^2 \widetilde{U}_0. \qquad (5.6)$$

Therefore, we connect our evolution equations

$$\widetilde{U}_t = JL_1^n \widetilde{U}_0, \quad n = 0,1,2,$$

with the constants of motion h_0, h_1, h_2 as follows:

$$\widetilde{U}_t = J(L_1^n \widetilde{U}_0) = J \begin{pmatrix} \frac{\delta}{\delta r} & h_n \\ \frac{\delta}{\delta q} & h_n \\ \frac{\delta}{\delta \beta} & h_n \\ \frac{\delta}{\delta \alpha} & h_n \end{pmatrix}, \quad \text{for } n = 0,1,2. \qquad (5.7)$$

This inspires us to prove that the above equations are also valid for all n, and are not limited to the first three, i.e.,

$$\widetilde{U}_t = \begin{pmatrix} r \\ q \\ \beta \\ \alpha \end{pmatrix}_t = J \begin{pmatrix} \frac{\delta}{\delta r} & h_{n-1} \\ \frac{\delta}{\delta q} & h_{n-1} \\ \frac{\delta}{\delta \beta} & h_{n-1} \\ \frac{\delta}{\delta \alpha} & h_{n-1} \end{pmatrix}$$

$$= \begin{pmatrix} \widetilde{B} \\ \widetilde{C} \\ 2\widetilde{\rho} \\ -2\widetilde{\delta} \end{pmatrix}, \quad n = 0,1,2,3,\ldots . \qquad (5.8)$$

Here,

$$h_{n-1} = qf_n + \alpha g_n. \qquad (5.9)$$

VI. CONSTRAINED VARIATION AND HAMILTONIAN STRUCTURE

We shall prove the Hamiltonian structure of the super AKNS scheme by using the constrained variational calculus.[5,6,8] First of all let us introduce the functional

$$H = h + \mu_1 X_1 + \mu_2 X_2 + \mu_3 X_3, \qquad (6.1)$$

where

$$h = qF + \alpha G = \frac{h_0}{\xi} + \frac{h_1}{\xi^2} + \cdots = \sum_{j=0}^{\infty} \frac{h_j}{\xi^{j+1}},$$

and for Lagrangian multipliers, μ_1, μ_2 are even numbers and μ_3 is odd, and

$$X_1 = h - (qF + \alpha G),$$
$$X_2 = -F_x + (2\xi F + r - qF^2 + \beta G - \alpha FG), \qquad (6.2)$$
$$X_3 = -G_x + (\xi G - \beta + \alpha F - qFG).$$

Here, as in Eq. (5.2) of Ref. 4,

$$F_x = 2\xi F + r - qF^2 + \beta G - \alpha FG,$$
$$G_x = \xi G - \beta + \alpha F - qFG. \qquad (6.3)$$

Using the constrained calculus for the super system, we get

$$\frac{\delta H}{\delta h} = 1 + \mu_1 = 0, \qquad (6.4a)$$

$$\frac{\delta H}{\delta F} = \mu_{2x} + 2\xi\mu_2 - 2qF\mu_2 - \alpha G\mu_2 + \mu_3\alpha - \mu_3 qG - \mu_1 q = 0, \qquad (6.4b)$$

$$\frac{\delta H}{\delta G} = -\beta\mu_2 + \alpha F\mu_2 - \mu_{3x} - \mu_3\xi + qF\mu_3 + \mu_1\alpha = 0, \qquad (6.4c)$$

and

$$\frac{\delta h}{\delta q} = -\mu_1 F - \mu_2 F^2 - \mu_3 FG, \qquad (6.5a)$$

$$\frac{\delta h}{\delta r} = \mu_2, \qquad (6.5b)$$

$$\frac{\delta h}{\delta \alpha} = -\mu_1 G - \mu_2 FG - \mu_3 F, \qquad (6.5c)$$

$$\frac{\delta h}{\delta \beta} = \mu_2 G + \mu_3. \qquad (6.5d)$$

We define

$$\widetilde{B} \equiv \frac{\delta}{\delta r} h, \quad \widetilde{C} \equiv \frac{\delta}{\delta q} h, \quad \widetilde{\delta} \equiv \frac{-1}{2} \frac{\delta}{\delta \alpha} h, \quad \widetilde{\rho} \equiv \frac{1}{2} \frac{\delta}{\delta \beta} h, \qquad (6.6)$$

and introduce

$$\widetilde{A} \equiv -\widetilde{B}F + \tfrac{1}{2} - \widetilde{\rho}G.$$

Then we would like to prove that $\widetilde{A}$, $\widetilde{B}$, $\widetilde{C}$, $\widetilde{\rho}$, and $\widetilde{\delta}$ also satisfy Eq. (3.4), as A, B, C, ρ, δ do.

From Eq. (6.4a), we obtain

$$\mu_1 = -1. \qquad (6.7)$$

From Eq. (6.5b) and $\widetilde{B} = \delta h / \delta r$, we have

$$\mu_2 = \widetilde{B}. \qquad (6.8)$$

From Eq. (6.5d) and $2\tilde{\rho} \equiv \delta h/\delta\beta$, we get

$$\mu_3 = 2\tilde{\rho} - \tilde{B}G. \tag{6.9}$$

On the other hand, from the definitions $\tilde{C} = \delta h/\delta q$, $-2\tilde{\delta} = \delta h/\delta\alpha$, and Eqs. (6.5a) and (6.5b) we get

$$\tilde{C} = F(1 - \tilde{B}F) - 2\tilde{\rho}FG \tag{6.10}$$

and

$$\tilde{\delta} = \tilde{\rho}F - \tfrac{1}{2}G, \tag{6.11}$$

respectively. Then, we must estimate the corresponding derivatives $\tilde{A}_x, \tilde{B}_x, \tilde{C}_x, \tilde{\rho}_x, \tilde{\delta}_x$, and show that they satisfy the same equations as $\bar{A}, \bar{B}, \bar{C}, \bar{\rho}, \bar{\delta}$ do.

A. Calculation of $\tilde{B}_x$

From Eq. (6.4b) it follows that

$$q + \tilde{B}_x + B(2\xi - 2qF - \alpha G) + (2\tilde{\rho} - \tilde{B}G)(\alpha - qG)$$
$$= \tilde{B}_x + 2\xi\tilde{B} + 2q\tilde{A} + 2\tilde{\rho}\alpha$$

and hence

$$\tilde{B}_x = -2\xi\tilde{B} - 2q\tilde{A} - 2\tilde{\rho}\alpha. \tag{6.12}$$

B. Calculation of $\tilde{\rho}_x$

From Eq. (6.4c) it follows that

$$\frac{\delta H}{\delta G} = \mu_1\alpha - \mu_2(\beta - \alpha F) - (\mu_{3x} + \mu_3\xi - \mu_3 qF) = 0,$$

and we have

$$-\alpha - \tilde{B}(\beta - \alpha F) - (2\rho - \tilde{B}G)_x - (2\tilde{\rho} - \tilde{B}G)\xi + (2\tilde{\rho} - \tilde{B}G)qF$$
$$= -\beta\tilde{B} + \alpha F\tilde{B} - 2\tilde{\rho}_x + \tilde{B}(\xi G - \beta + \alpha F - qFG) + (-2\xi\tilde{B} + 2qF\tilde{B}$$
$$- 2\tilde{\rho}\alpha + 2\tilde{\rho}qG - q)G - 2\tilde{\rho}\xi + \tilde{B}G\xi + 2qF\tilde{\rho} - qF\tilde{B}G - \alpha = 0.$$

Notice $\tilde{\delta} = \tilde{\rho}F - \frac{1}{2}G$, we get

$$-\tilde{\rho}_x - \tilde{\rho}\xi + q\tilde{\delta} - \beta\tilde{B} + \alpha F\tilde{B} - \tilde{\rho}\alpha G - \tfrac{1}{2}\alpha,$$

i.e.,

$$\tilde{\rho}_x = -\tilde{\rho}\xi + q\tilde{\delta} - \tilde{B}\beta - \alpha\tilde{A}. \tag{6.13}$$

C. Calculation of $\tilde{\delta}_x$

From Eq. (6.11), we have

$$\tilde{\delta}_x = \tilde{\rho}_x F + \tilde{\rho}F_x - \tfrac{1}{2}G_x$$
$$= (-\tilde{\rho}\xi + q\tilde{\delta} - \beta\tilde{B} + \alpha F\tilde{B} - \tilde{\rho}\alpha G - \tfrac{1}{2}\alpha)F + 2\tilde{\rho}(2\xi F + r - qF^2 + \beta G - \alpha GF) - (-\xi G + \beta - \alpha F + qFG)$$
$$= \xi\tilde{\delta} - \beta\tilde{A} + r\tilde{\rho} - \alpha\tilde{C}. \tag{6.14}$$

D. Calculation of $\tilde{A}_x$

$$\tilde{A}_x = (-\tilde{B}F + \tfrac{1}{2} - \tilde{\rho}G)_x = -(\tilde{B}_x F + \tilde{B}F_x + \tilde{\rho}_x G + \tilde{\rho}G_x).$$

Using F_x, G_x in (6.3), $\tilde{B}_x, \tilde{\rho}_x$ in (6.12), (6.13), and after tedious recapitulation of the terms, we get

$$\tilde{A}_x = q\tilde{C} - r\tilde{B} - \alpha\tilde{\delta} - \beta\tilde{\rho}. \tag{6.15}$$

E. Calculation of C_x

$$\tilde{C}_x = F_x - 2\tilde{B}FF_x - \tilde{B}_x F^2 - 2(\tilde{\rho}FG)_x$$
$$= [2\xi F(1 - \tilde{B}F - 2\tilde{\rho}G) - 2r(\tilde{B}F + \tilde{\rho}G) + r] + (-qF^2 + \beta G - \alpha GF - 4\xi\tilde{B}F^2 + 2\tilde{B}qF^3 + 2\tilde{B}F^2\alpha G$$
$$+ 2q\tilde{A}F^2 + 2\tilde{\rho}\alpha F^2 + 2q\tilde{\rho}F^2G + 2\alpha\tilde{A}FG - 2\tilde{\rho}\alpha F^2 + 2\tilde{\rho}\beta F).$$

After extremely laborious computation and skill, we obtain

$$\tilde{C}_x = 2(\xi\tilde{C} + r\tilde{A} - \beta\delta). \tag{6.16}$$

To sum up, the functions $\tilde{A},\tilde{B},\tilde{C},\tilde{\rho},\tilde{\delta}$ and $\bar{A},\bar{B},\bar{C},\bar{\rho},\bar{\delta}$ satisfy the same linear equation (2.8). From (6.6) and (2.9), we find

$$a_n = -2\tilde{a}_n, \quad b_{n+1} = -2\frac{\delta h_n}{\delta r}, \quad c_{n+1} = -2\frac{\delta h_n}{\delta^q}, \tag{6.17}$$

$$\rho_{n+1} = -\frac{\delta h_n}{\delta\beta}, \quad \delta_{n+1} = \frac{\delta h_n}{\delta\alpha},$$

using (4.4), we conclude

$$\widetilde{U}_t = JL_1^n \widetilde{U}_0 = J \begin{pmatrix} \frac{\delta h_n}{\delta r} \\ \frac{\delta h_n}{\delta q} \\ \frac{\delta h_n}{\delta \beta} \\ \frac{\delta h_n}{\delta \alpha} \end{pmatrix}. \tag{6.18}$$

This is the super-Hamiltonian form of the evolution equation (of motion), where the supersymplectic operator J had the form as expressed in Eq. (4.2).

ACKNOWLEDGMENTS

The authors would like to express their sincere thanks to the referee for his helpful comments.

[1]B. A. Kupershmidt, Phys. Lett. A **102**, 213 (1984).
[2]M. Gurses and O. Oguz, Phys. Lett. A **108**, 437 (1985).
[3]M. Omote and K. Inono, Prov. Theor. Phys. **72**, 641 (1984).
[4]A. R. Choudhury and S. Roy, J. Math. Phys. **27**, 2464 (1986).
[5]Yun-bo Zheng, Yi-shen Li, and Dong-Yuan Chen, Sci. Sinica A **24**, 138 (1986).
[6]Gui-zhang Tu and Kexue Tongbo (Science Building) **29**, 1227 (1984).
[7]Li Ti-shen and Zhang Li-ning, Nuovo Cimento A **93**, 175 (1986).
[8]Yi-shen Li and A. Degasperis, Super Kdv hierarchy, private communication (1985).

A superparticle on the super Riemann surface

Shuji Matsumoto
Department of Physics, Osaka City University, Sumiyoshi, Osaka 558, Japan
Shozo Uehara
Research Institute for Theoretical Physics, Hiroshima University, Takehara, Hiroshima 725, Japan
Yukinori Yasui
Department of Physics, Osaka City University, Sumiyoshi, Osaka 558, Japan

(Received 18 July 1989; accepted for publication 20 September 1989)

The free motion of a nonrelativistic superparticle on the super Riemann surface (SRS) of genus $h \geqslant 2$ is investigated. Geodesics or classical paths are given explicitly on the super Poincaré upper half-plane SH, a universal covering space of the SRS, and the paths with some suitable initial conditions yield periodic orbits on the SRS. The periodic orbits are unstable and the system is chaotic. Quantum mechanics is solved on the universal covering space SH and the heat kernel is given on the SRS. This leads to a superanalog of the Selberg trace formula. The Selberg super zeta function is introduced whose zero points and poles determine the energy spectrum on the SRS.

I. INTRODUCTION

The purpose of this paper is to examine a supersymmetric extension of the Hadamard model[1] that represents the free motion of a nonrelativistic particle on a compact Riemann surface of constant negative curvature. That is, we investigate the system of a superparticle moving freely on a compact super Riemann surface (SRS) of genus $h \geqslant 2$. The former is known as one of the typical models of chaotic systems.[2] The Riemann surface is represented by a fundamental domain H/Γ in the complex upper half-plane H, the universal covering space of the Riemann surface, with Γ being a Fuchsian group that is a discrete subgroup of PSL(2,$\mathbb{R}$). The Γ-invariant Lagrangian is made of the line element $ds_0^2 = |dz|^2/(\operatorname{Im} z)^2$,

$$L_0 = \frac{m}{2}\left(\frac{ds_0}{dt}\right)^2. \tag{1}$$

The classical motions on H are integrable, however, those on the compact Riemann surface are chaotic and the energy is the only conserved quantity. The quantized energy sum rule is actually the Selberg trace formula.[3,4] The energy spectrum is complicated, however, it is given by examining the Selberg zeta function. The Selberg trace formula or zeta function appears in the Polyakov partition function for a closed bosonic string.[5] The notion of the super Riemann surfaces comes naturally in the superspace approach of superstrings.[6] Superanalogs of the trace formula and the zeta function are important for the theory.

Some part of the results here will be seen in our recent papers,[7-10] however, for self-containedness, we shall present those results here again. This paper is organized as follows. Section II is devoted to the notations and conventions of the basic facts on super Riemann surfaces. We give a super Möbius-invariant Lagrangian for a superparticle on a super Riemann surface of genus $h \geqslant 2$. Classical mechanics for the system is developed in Sec. III. We discuss chaos in this system there. Quantization is carried out in Sec. IV. The eigenvalue problem in the universal covering space SH is developed there and the kernel function is given. In Sec. V a superanalog of the Selberg trace formula is given and a superanalog of the Selberg zeta function is introduced. The final section is devoted to summary and discussions. Some of the detailed calculations are presented in the Appendices.

II. PRELIMINARIES

A. Super Riemann surfaces

A super Riemann surface having a compact body with $h \geqslant 2$ holes is represented by a homogeneous space SH/SΓ[11,12] with a superanalog of the Poincaré geometry. The universal covering space of the SRS is the super complex upper half-plane SH with one even and one odd complex coordinate z and θ, respectively.

$$\mathrm{SH} = \{Z = (z,\theta) \mid \operatorname{Im} z > 0\}. \tag{2}$$

Note that $\operatorname{Im} z > 0$ means that $\operatorname{Im} z_0 > 0$ with z_0 being the body part of z. We shall use such a convention for inequalities throughout this paper for simplicity. SΓ is called a super Fuchsian group that is a discrete subgroup of superconformal automorphisms SPL(2,$\mathbb{R}$) of SH. The supergroup SPL(2,R) consists of such transformations as,

$$\begin{aligned} z \to \tilde{z} &= \frac{az+b}{cz+d} + \theta\,\frac{\alpha z + \beta}{(cz+d)^2},\\ \theta \to \tilde{\theta} &= \frac{\alpha z+\beta}{cz+d} + \theta\,\frac{1+\frac{1}{2}\beta\alpha}{cz+d}, \end{aligned} \tag{3}$$

where a, b, c, and d are Grassmann even and α and β are Grassmann odd parameters with[13]

$$\begin{aligned} &ad - bc = 1, \quad a,b,c,d \in \mathbb{R},\\ &\bar{\alpha} = i\alpha, \qquad \bar{\beta} = i\beta. \end{aligned} \tag{4}$$

Note that the above transformation (3) is, of course, superanalytic and is also a superconformal transformation,

$$D\tilde{z} - \tilde{\theta} D\tilde{\theta} = 0, \tag{5}$$

$$D \equiv \frac{\partial}{\partial\theta} + \theta\frac{\partial}{\partial z}. \tag{6}$$

 0022-2488/90/020476-26$03.00

If we introduce homogeneous coordinates (z_1,z_2,ξ) of complex projective lines, we can rewrite (3) as a linear transformation with $z=z_1z_2^{-1}$, $\theta=\xi z_2^{-1}$,

$$\begin{pmatrix}\tilde z_1\\ \tilde z_2\\ \tilde\xi\end{pmatrix}=A\begin{pmatrix}z_1\\ z_2\\ \xi\end{pmatrix},$$

$$A=(1+\tfrac{1}{2}\beta\alpha)^{-1}\times\begin{pmatrix}a & b & b\alpha-a\beta\\ c & d & d\alpha-c\beta\\ \alpha & \beta & 1+\tfrac{3}{2}\beta\alpha\end{pmatrix},\quad \mathrm{sdet}\,A=1. \tag{7}$$

The super Fuchsian group $S\Gamma$ is generated by $2h$ elements $\{A_i,B_i,\ i=1,...,h\}$ that satisfy the condition

$$\prod_{i=1}^{h}(A_iB_iA_i^{-1}B_i^{-1})=1. \tag{8}$$

$S\Gamma$ properly acts discontinuously on SH and all its elements are hyperbolic, i.e., the reduced subgroup, where odd parameters are put to zero, consists of the hyperbolic elements, $|a+d|>2$.

B. Conjugacy classes in $S\Gamma$

An element $k\neq 1$ of $S\Gamma$ causes such a transformation as (3) with (4). $S\Gamma$ acts effectively on SH, however, $k\in S\Gamma$ has fixed points on the "super" real axis, $\mathbb{R}_s\equiv\{Z=(z,\theta)\,|\,\mathrm{Im}\,z=0,\ \bar\theta=i\theta\}$. In fact, the fixed points $(u_\pm,\mu_\pm)$ are given by

$$u_\pm=\frac{a-d\pm\sqrt{(a+d)^2-4}}{2c},\quad \mu_\pm=\frac{\alpha u_\pm+\beta}{cu_\pm+d-1}. \tag{9}$$

We can easily see that $\bar u_\pm=u_\pm$, $\bar\mu_\pm=i\mu_\pm$.[14] Using the fixed points, $(u_\pm,\mu_\pm)=(u,\mu)$ and (v,ν), we can rewrite the transformation as a magnification,

$$w\to\tilde w=Nw,$$
$$\eta\to\tilde\eta=\chi N^{1/2}\eta,\quad N>1, \tag{10}$$

where

$$\chi(N^{1/2}+N^{-1/2})=a+d-[(a+d+2)/2]\beta\alpha$$
$$=\mathrm{str}\,A+1,$$
$$\chi=\begin{cases}1, & \text{if str}\,A+1>2,\\ -1, & \text{if str}\,A+1<-2,\end{cases} \tag{11}$$

$$w=\left(\frac{1}{u-v}\right)\frac{z-u-\theta\mu}{z-v-\theta\nu},$$
$$\eta=\left(\frac{1}{u-v}\right)\frac{(u-v+\tfrac{1}{2}\nu\mu)\theta+(\nu-\mu)z+\mu v-\nu u}{z-v+\nu\theta}. \tag{12}$$

It is easy to see that the two fixed points (u,μ) and (v,ν) are repelling and attractive points, respectively,

$$k^{-n}:\ (z,\theta)\to(u,\mu),$$
$$k^{\ n}:\ (z,\theta)\to(v,\nu),\quad \text{for } n\to\infty. \tag{13}$$

Since the transformation $f:(z,\theta)\to(w,\eta)$ in (12) is an element of SPL(2,$\mathbb{R}$), we see that any element $\neq 1$ of $S\Gamma$ is conjugate in SPL(2,$\mathbb{R}$) to magnification,

$$A_fA_kA_f^{-1}=A_{fkf^{-1}}$$
$$=\begin{pmatrix}\chi(k)N_k^{1/2} & 0 & 0\\ 0 & \chi(k)N_k^{-1/2} & 0\\ 0 & 0 & 1\end{pmatrix}\equiv A_{\mathrm{mag}}. \tag{14}$$

Apparently, the magnification depends on the conjugacy class of k in $S\Gamma$, $\{k\}=\{gkg^{-1}|g\in S\Gamma\}$,

$$N_{gkg^{-1}}=N_k,\quad \chi(gkg^{-1})=\chi(k). \tag{15}$$

C. Decomposition of $S\Gamma$

The element $p\neq 1$ is called "primitive" if it is not a power of any other element of $S\Gamma$. The centralizer $Z(q)$ of an element q in $S\Gamma$ is given by

$$Z(q)\equiv\{g|gqg^{-1}=q,\ g\in S\Gamma\}. \tag{16}$$

Due to the fact that

$$AA_{\mathrm{mag}}=A_{\mathrm{mag}}A\Leftrightarrow A=\begin{pmatrix}a & 0 & 0\\ 0 & a^{-1} & 0\\ 0 & 0 & 1\end{pmatrix}, \tag{17}$$

with A and A_{mag} being given in (7) and (14), respectively, we see that if $k\neq 1$ is a magnification, then any element of the centralizer of the k is a magnification. This implies that there exists such a primitive element p that k is expressed uniquely as a positive power of the element p and hence the centralizer $Z(k)$ is, in fact, the cyclic group $\langle p\rangle$ generated by the primitive element p.

Let us consider the following set;

$$Q\equiv\{gp^ng^{-1}|g\in S\Gamma/Z(p)\}. \tag{18}$$

The above arguments lead that (i) Q runs once through the nontrivial conjugacy classes of $S\Gamma$ when p runs through inconjugate primitive elements of $S\Gamma$ and n through positive integers and (ii) for fixed p and n, qp^nq^{-1} runs once through Q as q runs through $S\Gamma/Z(p)$. Then for a function f of elements of $S\Gamma$, we get

$$\sum_{g\in S\Gamma}f(g)=f(1)+\sum_{\substack{\text{inconjugate}\\ \text{primitive } p}}\sum_{n=1}^{\infty}\sum_{g\in S\Gamma/Z(p)}f(gp^ng^{-1}). \tag{19}$$

D. Metric on SH

In this subsection we will introduce a SPL(2,$\mathbb{R}$)-invariant metric on SH that is a superanalog of the Poincaré metric on H. The latter,

$$ds_0^2=|dz|^2/(\mathrm{Im}\,z)^2, \tag{20}$$

is invariant under PSL(2,$\mathbb{R}$) and gives a constant negative curvature $R=-2$. The corresponding volume element is

$$(dx\,dy)/y^2\quad(\mathrm{Re}\,z=x,\ \mathrm{Im}\,z=y). \tag{21}$$

For the purpose of giving a SPL(2,$\mathbb{R}$)-invariant metric, we first comment on SPL(2,$\mathbb{R}$)-covariant quantities. Let Z_a, Z_b, and Z_c be three points in superspace. Then the following quantities are SPL(2,$\mathbb{R}$)-covariant[15]:

$$z_{ab} = z_a - z_b - \theta_a \theta_b$$
$$z_{\bar{a}b} = z_{\bar{a}} - z_b - \theta_{\bar{a}} \theta_b, \ldots, \text{etc.},$$
$$\theta_{abc} = \theta_a z_{bc} + \theta_b z_{ca} + \theta_c z_{ab} + \theta_a \theta_b \theta_c,$$
$$\theta_{\bar{a}bc} = \theta_{\bar{a}} z_{bc} + \theta_b z_{c\bar{a}} + \theta_c z_{\bar{a}b} + \theta_{\bar{a}} \theta_b \theta_c, \ldots, \text{etc.}, \quad (22)$$

where coordinates with barred suffixes are defined by

$$z_{\bar{a}} \equiv \overline{z_a}, \quad \theta_{\bar{a}} \equiv -i\, \overline{\theta_a}, \ldots, \text{etc.}, \quad (23)$$

and hence

$$z_{\bar{a}\bar{b}} = \overline{z_{ab}}, \quad \theta_{\bar{a}\bar{b}\bar{c}} = -i\, \overline{\theta_{abc}}, \ldots, \text{etc.} \quad (24)$$

In fact, we see

$$\tilde{z}_{ab} \equiv \tilde{z}_a - \tilde{z}_b - \tilde{\theta}_a \tilde{\theta}_b = \Omega_a \Omega_b z_{ab}, \quad \tilde{z}_{\bar{a}b} = \Omega_{\bar{a}} \Omega_b z_{\bar{a}b}, \ldots, \text{etc.},$$
$$\tilde{\theta}_{abc} = \Omega_a \Omega_b \Omega_c \theta_{abc}, \quad \tilde{\theta}_{\bar{a}bc} = \Omega_{\bar{a}} \Omega_b \Omega_c \theta_{\bar{a}bc}, \ldots, \text{etc.}, \quad (25)$$

where the transition function Ω is

$$\Omega_a = (D\tilde{\theta})_a, \quad \Omega_{\bar{a}} = \overline{(\Omega_a)}, \ldots, \text{etc.} \quad (26)$$

Taking $Z_1 = (z,\theta)$, and its infinitesimally neighboring point $Z_2 = (z + dz,\ \theta + d\theta)$, we have the following covariant quantities:

$$z_{21} = dz + \theta\, d\theta, \quad (27)$$
$$(2i)^{-1} z_{1\bar{1}} = \operatorname{Im} z + \tfrac{1}{2} \theta \bar{\theta} \equiv Y, \quad (28)$$

which are superanalogs of dz and y, respectively;

$$d\tilde{Z} = \Omega^2\, dZ, \quad \tilde{Y} = |\Omega|^2 Y. \quad (29)$$

We also see that

$$|dz + \theta\, d\theta|^2 / Y^2 \quad (30)$$

is a SPL(2,$\mathbb{R}$)-invariant quantity whose body part is (21).

On the other hand, the geometry of a 2 + 2-dimensional superspace was developed in supergravity theory. The basic quantities are the super vielbein E_M^A which, however, are not completely independent superfields. It was shown that 2 + 2-dimensional superspace is superconformally flat[16] where the basis one-forms $\hat{E}^A$ are

$$\hat{E}^{++} = dz + \theta\, d\theta, \quad \hat{E}^{--} = d\bar{z} - \bar{\theta}\, d\bar{\theta},$$
$$\hat{E}^{+} = d\theta, \quad \hat{E}^{-} = d\bar{\theta}. \quad (31)$$

By the super Weyl transformation[16] with the parameter Y^{-1}, $\hat{E}^A$ become

$$E^{++} = Y^{-1} \hat{E}^{++} = Y^{-1}(dz + \theta\, d\theta), \quad (32)$$

$$E^{--} = Y^{-1}(d\bar{z} - \bar{\theta}\, d\bar{\theta}) (= \overline{E^{++}}), \quad (33)$$

$$E^{+} = Y^{-1/2} \hat{E}^{+} + 2(DY^{-1/2}) \hat{E}^{++}$$
$$= Y^{-3/2}[(Y + \tfrac{1}{2}\theta\bar{\theta}) d\theta + \tfrac{1}{2}(i\theta - \bar{\theta}) dz], \quad (34)$$

$$E^{-} = Y^{-3/2}[(Y + \tfrac{1}{2}\theta\bar{\theta}) d\bar{\theta} - \tfrac{1}{2}(\theta + i\bar{\theta}) d\bar{z}] (= \overline{E^{+}}). \quad (35)$$

We see that (30) is $(E^{++}E^{--})$ and moreover, taking $Z_1 = (z,\theta)$ and $Z_2 = (z + dz, \theta + d\theta)$ as before, we find that

$$\theta_{21\bar{1}} = 2i[(Y + \tfrac{1}{2}\theta\bar{\theta}) d\theta + \tfrac{1}{2}(i\theta - \bar{\theta}) dz]$$
$$= (2i)^{-1/2} (z_{1\bar{1}})^{3/2} E^{+}, \quad (36)$$

and hence

$$\tilde{E}^{+} = \Omega^{1/2} \bar{\Omega}^{-1/2} E^{+}. \quad (37)$$

Then, $(E^{+}E^{-})$ is also SPL(2,$\mathbb{R}$)-invariant, which is nilpotent and vanishes when the odd coordinate θ is put to zero.

We now introduce a SPL(2,R)-invariant metric on SH given by[10,17]

$$ds^2 = E^{++}E^{--} - 2aE^{+}E^{-},$$
$$\equiv dq^A g_{AB}\, dq^B, \quad (q^z, q^{\bar{z}}, q^{\theta}, q^{\bar{\theta}}) = (z, \bar{z}, \theta, \bar{\theta}), \quad (38)$$

$$(g_{AB}) = \begin{pmatrix} 0 & \dfrac{1}{2Y^2} & 0 & -\dfrac{\bar{\theta} + a(i\theta - \bar{\theta})}{2Y^2} \\ \dfrac{1}{2Y^2} & 0 & \dfrac{\theta - a(\theta + i\bar{\theta})}{2Y^2} & 0 \\ 0 & \dfrac{a(\theta + i\bar{\theta}) - \theta}{2Y^2} & 0 & \dfrac{\theta\bar{\theta} - 2a(Y + \theta\bar{\theta})}{2Y^2} \\ \dfrac{\bar{\theta} + a(i\theta - \bar{\theta})}{2Y^2} & 0 & \dfrac{2a(Y + \theta\bar{\theta}) - \theta\bar{\theta}}{2Y^2} & 0 \end{pmatrix}$$

(a: arbitrary real Grassmann even number$\neq 0$). (39)

Geometrical quantities such as the Riemann tensor, the Ricci tensor, etc. are given in Appendix A. The corresponding volume element is given by

$$dV = (1/2aY) dx\, dy\, d\theta\, d\bar{\theta}. \quad (40)$$

E. Lagrangian

Since we have now a SPL(2,R)-invariant line element (38), we give the Lagrangain of a superparticle with mass m on SH,[10]

$$L = \frac{m}{2}\left(\frac{ds}{dt}\right)^2 = \frac{m}{2} \dot{q}^A g_{AB} \dot{q}^B. \quad (41)$$

This is SPL(2,R)-invariant and hence, of course, SΓ-invariant, thus it is also the Lagrangian for a superparticle on the SRS.

III. CLASSICAL MECHANICS

In this section we examine the classical dynamics of a superparticle on the SRS. We first develop the canonical theory with the Lagrangian (41). The canonical momentum is given by

$$p_A = \frac{\partial L}{\partial \dot{q}^A} = m g_{AB} \dot{q}^B \tag{42}$$

or

$$\begin{aligned} p_z &= (m/2Y^2)[\dot{\bar{z}} + \{\bar{\theta} + ia(\theta + i\bar{\theta})\}\dot{\bar{\theta}}], \\ p_\theta &= (m/2Y^2)[\{a(\theta + i\bar{\theta}) - \theta\}\dot{\bar{z}} \\ &\quad + \{\theta\bar{\theta} - 2a(Y + \theta\bar{\theta})\}\dot{\bar{\theta}}], \\ p_{\bar{z}} &= \overline{(p_z)}, \quad p_{\bar{\theta}} = - \overline{p_\theta}, \end{aligned} \tag{43}$$

with the Hamiltonian given by

$$H = \dot{q}^A p_A - L = (1/2m) g^{BA} p_A p_B, \tag{44}$$

where g^{AB} is the inverse metric to g_{AB} (see Appendix A). The Poisson bracket is defined by,[18,19]

$$\{F,G\}_P \equiv (-)^{AF} \frac{\partial F}{\partial q^A} \frac{\partial G}{\partial p_A} - (-)^{A(F+1)} \frac{\partial F}{\partial p_A} \frac{\partial G}{\partial q^A}, \tag{45}$$

so that

$$\{p_A, q^B\}_P = -\delta_A^B, \tag{46}$$

and the Hamilton equations are

$$\{q^A, H\}_P = \dot{q}^A, \quad \{p_A, H\}_P = \dot{p}_A. \tag{47}$$

The Lagrangian (41) is SPL(2,R)-invariant. The conserved charges are given by

$$\begin{aligned} L_{-1} &= p_z + p_{\bar{z}}, \\ L_0 &= z p_z + \bar{z} p_{\bar{z}} + \tfrac{1}{2}\theta p_\theta + \tfrac{1}{2}\bar{\theta} p_{\bar{\theta}}, \\ L_1 &= z^2 p_Z + \bar{z}^2 p_{\bar{z}} + z\theta p_\theta + \bar{z}\bar{\theta} p_{\bar{\theta}}, \\ G_{-1/2} &= \theta p_z - i\bar{\theta} p_{\bar{z}} - p_\theta - i p_{\bar{\theta}}, \\ G_{1/2} &= z\theta p_z - i\bar{z}\bar{\theta} p_{\bar{z}} - z p_\theta - i\bar{z} p_{\bar{\theta}}. \end{aligned} \tag{48}$$

The Poisson brackets between them yield,

$$\begin{aligned} \{L_m, L_n\}_P &= (m-n) L_{m+n}, \\ \{L_m, G_r\}_P &= (m/2 - r) G_{m+r}, \\ \{G_r, G_s\}_P &= 2L_{r+s} \quad (m,n = 0, \pm 1, r,s = \pm \tfrac{1}{2}). \end{aligned} \tag{49}$$

Next, we study the classical motion on the covering space SH. The Euler–Lagrange equations from L in (41) are geodesic equations,

$$\ddot{q}^A + \Gamma^A_{BC} \dot{q}^C \dot{q}^B = 0, \tag{50}$$

where the Cristoffel's symbol is given in Appendix A, or more explicitly,[10]

$$\begin{aligned} &\ddot{z} + \frac{1}{Y}(i\dot{z}^2 - \dot{z}\theta\bar{\theta}) + \frac{1-a}{2a}\left(\frac{i}{Y^2}\theta\bar{\theta}\dot{z}\dot{\bar{z}} - \frac{2}{Y}\dot{z}\theta\dot{\bar{\theta}}\right) = 0, \\ &\ddot{\theta} + \frac{i}{Y}\dot{z}\dot{\theta} + \frac{1-a}{2a}\left(\frac{2Y + \theta\bar{\theta}}{Y^2}\dot{z}\dot{\bar{\theta}}\right. \\ &\quad \left. - \frac{\theta + i\bar{\theta}}{Y^2}\dot{z}\dot{\bar{z}} + \frac{2}{Y}\theta\dot{\theta}\dot{\bar{\theta}} + \frac{i}{Y^2}\dot{\bar{z}}\theta\bar{\theta}\dot{\theta}\right) = 0, \end{aligned} \tag{51}$$

and their complex conjugated ones. The body part of Eqs. (51) is

$$\ddot{z}_0 + (i/y_0)\dot{z}_0^2 = 0, \tag{52}$$

which is the geodesic equation on H with the Poincaré metric.[1] The solutions of (52) are given by,[20]

$$z_0(t) = c_1 \frac{\sinh X_0 + i}{\cosh X_0} + c_2, \quad \text{or} \quad ie^{X_0} + c_2, \tag{53}$$

where

$$X_0 \equiv \omega(t + t_0), \quad c_1, c_2, \omega, t_0 \in \mathbb{R}. \tag{54}$$

The classical motion is determined uniquely with the boundary conditions that are positions and velocities at the initial point. Thus the constants of the integration for the Euler–Lagrange equation (50) or (51) are four real Grassmann even and also four odd constants. Expanding z and θ in the Grassmann odd constants, say, $\epsilon_1, \bar{\epsilon}_1, \epsilon_2, \bar{\epsilon}_2$, we have a set of differential equations for the coefficients of the Grassmann even functions. However, it is not easy to solve those equations. So instead of solving (51) directly, we will take a roundabout route. Due to (44), the Hamilton–Jacobi equation is given by

$$\frac{\partial S}{\partial t} + \frac{1}{2m} g^{BA} \frac{\partial S}{\partial q^A} \frac{\partial S}{\partial q^B} = 0. \tag{55}$$

Since the action which the classicl solutions are plugged into satisfies the above equation (55), we express S as

$$S(q_1, q_2; t_2 - t_1) = \frac{m}{2} \int_{t_1}^{t_2} dt\, \dot{q}^A(t) g_{AB}(q(t)) \dot{q}^B(t), \tag{56}$$

where $q^A(t)$ is a solution of the geodesic equation (50) connecting the initial point $q_i = q(t_1)$ and the final point $q_f = q(t_2)$. It can be easily shown that the integrand is time-independent and its body part is non-negative. Taking them into account, we set

$$\dot{q}^A(t) g_{AB}(t) \dot{q}^B(t) = (\text{const}) \equiv \omega^2, \tag{57}$$

and define a superanalog of the hyperbolic distance,

$$\begin{aligned} d(q_1, q_2) &= \int_{t_1}^{t_2} dt \sqrt{\dot{q}^A(t) g_{AB}(t) \dot{q}^B(t)} \\ &= |\omega| (t_2 - t_1). \end{aligned} \tag{58}$$

From (56), (57), and (58), we get

$$S(q_1, q_2; t) = (m/2)([d(q_1, q_2)]^2/t). \tag{59}$$

Note that the hyperbolic distance $d_0(q_1, q_2)$ between $((q_1)_0) = (z_0, \bar{z}_0)$ and $((q_2)_0) = (w_0, \bar{w}_0)$, which should be the body part of $d(q_1, q_2)$, is given by

$$\cosh d_0 = 1 + \frac{|z_0 - w_0|^2}{2\, \text{Im}\, z_0\, \text{Im}\, w_0} \equiv 1 + \tfrac{1}{2} R_0, \tag{60}$$

and is PSL(2,R)-invariant. Hence $d(q_1, q_2)$ should by symmetric under the exchange of q_1 and q_2 and be SPL(2,R)-invariant. There exist two basic functions on SH×SH with such properties,[21]

$$R(q_1, q_2) \equiv 4 \frac{z_{12} z_{\bar{2}\bar{1}}}{z_{1\bar{1}} z_{2\bar{2}}} = \frac{|z_1 - z_2 - \theta_1\theta_2|^2}{(Y_{(1)} Y_{(2)})}, \tag{61}$$

$$r(q_1,q_2) \equiv \frac{\theta_{1\bar 2}\theta_{\bar 1\bar 2}}{z_{1\bar 1}(z_{2\bar 2})^2} = \frac{\theta_{2 1\bar 1}\theta_{\bar 2\bar 1 1}}{(z_{1\bar 1})^2 z_{2\bar 2}}$$
$$= \left\{ \frac{2\theta_1\bar\theta_1 + i(\theta_2 - i\bar\theta_2)(\theta_1 + i\bar\theta_1)}{4Y_{(1)}} + (1\leftrightarrow 2) \right\} + \frac{(\theta_2 + i\bar\theta_2)(\theta_1 + i\bar\theta_1)\mathrm{Re}(z_1 - z_2 - \theta_1\theta_2)}{4Y_{(1)}Y_{(2)}}. \tag{62}$$

Here, R is the superanalog of R_0 in (60) and r is nilpotent, and hence we can expect that, in general, $d(q_1q_2)$ takes the following form:

$$\cosh d = f(R) + k(R)r. \tag{63}$$

The calculation for determining the so far unknown functions f and k is given in Appendix B. We find that the "super" hyperbolic distance $d(q_1,q_2)$ is given by[10]

$$\cosh[d(q_1,q_2)] = 1 + \tfrac{1}{2}R(q_1,q_2) + k(R)r(q_1,q_2), \tag{64}$$

where

$$k(R) = \cosh(l) - 1 - \sinh(l)\coth(l/2a),$$
$$l = l(q_1,q_2) = \cosh^{-1}(1 + \tfrac{1}{2}R). \tag{65}$$

Here we have the solution (59) of the Hamilton–Jacobi equation.

The next step is to solve $q_1 \equiv q$ in terms of q_2 and its canonical conjugated quantity, say, $p^{(2)}$. This can be done by solving the following algebraic equations with respect to q:

$$\frac{\partial S}{\partial q_2^A} = -p_A^{(2)}, \tag{66}$$

where the q_2's and $p^{(2)}$'s actually correspond to the constants of integration for the differential equations (51),

$$q = q(q_2,p^{(2)},t). \tag{67}$$

The calculation is cumbersome but rather straightforward, which is shown in Appendix C. The solution of the Euler–Lagrange equations (51) is $(z^{(\mathrm{I})}(t),\theta^{(\mathrm{I})}(t))$, $(z^{(\mathrm{II})}(t),\theta^{(\mathrm{II})}(t))$ or $(z^{(\mathrm{III})}(t),\theta^{(\mathrm{III})}(t))$,[22]

$$z^{(\mathrm{I})}(t) = \left[c_1 - \frac{2}{\cosh X}\{i\xi_1\xi_2 e^{-X/a} - i\xi_3\xi_4 e^{X/a} - \xi_1\xi_4 e^{(1-1/a)X} + \xi_2\xi_3 e^{(1/a-1)X}\}\right] \times \frac{\sinh X + i}{\cosh X} + c_2, \tag{68}$$
$$\theta^{(\mathrm{I})}(t) = \left(\frac{\sinh X + i}{\cosh X} + 1\right)\{\xi_1 e^{-X/a} - i\xi_2 e^{-X} + i\xi_3 e^{(1/a-1)X} + \xi_4\},$$

$$z^{(\mathrm{II})}(t) = ie^X + c_2 + i\xi_1\xi_2 e^{(1-1/a)X} - i\xi_3\xi_4 e^{(1+1/a)X} - \xi_1\xi_4 e^{(2-1/a)X} + \xi_2\xi_3 e^{X/a}, \tag{69}$$
$$\theta^{(\mathrm{II})}(t) = i\xi_1 e^{(1-1/a)X} + \xi_2 - \xi_3 e^{X/a} + i\xi_4 e^X,$$

$$z^{(\mathrm{III})}(t) = ic_1 + c_2 - 2ac_1\omega_s t - \omega_s\{2ia^2c_1\omega_s - a\bar\epsilon_2\epsilon_1 + (1-a)\epsilon_2\bar\epsilon_1\}t^2 - \tfrac{1}{3}\epsilon_1\bar\epsilon_1\omega_s t^3, \tag{70}$$
$$\theta^{(\mathrm{III})}(t) = \epsilon_2 + \epsilon_1 t + [\{ia\epsilon_1 + (1-a)\bar\epsilon_1\}\omega_s - [(1-a)/(2ac_1)]\epsilon_1\bar\epsilon_1\epsilon_2]t^2,$$

where

$$X \equiv \omega(t + t_0), \tag{71}$$

and

ω, t_0, c_1, c_2: real Grassmann even constants, $c_1 > 0$,

$\xi_k\ (k = 1,2,3,4)$: Grassmann odd constants with $\bar\xi_k = i\xi_k$, (72)

ϵ_1, ϵ_2: complex Grassmann odd constants,

ω_s: Grassmann even constant with no body part,

that is, ω_s can be written by,

$$\omega_s = f_1\epsilon_1\bar\epsilon_1 + f_2\epsilon_1\epsilon_2 + f_3\epsilon_1\bar\epsilon_2 + f_4\bar\epsilon_1\epsilon_2 + f_5\bar\epsilon_1\bar\epsilon_2 + f_6\epsilon_2\bar\epsilon_2 + f_7\epsilon_1\bar\epsilon_1\epsilon_2\bar\epsilon_2, \tag{73}$$

with the f's being arbitrary complex constants. The first $(z^{(\mathrm{I})},\theta^{(\mathrm{I})})$ and the second $(z^{(\mathrm{II})},\theta^{(\mathrm{II})})$ solutions correspond to the first and the second solutions in (53), respectively with the third $(z^{(\mathrm{III})},\theta^{(\mathrm{III})})$ corresponding to the solution with $\omega = 0$ in (53). Actually, $(z^{(\mathrm{II})},\theta^{(\mathrm{II})})$ is obtained by taking a proper limit of $(z^{(\mathrm{I})},\theta^{(\mathrm{I})})$ (see Appendix C).

We now examine the classical motion on the SRS. Since we have obtained the classical paths (68), (69), (70) on the universal covering space SH of the SRS, we can deduce the classical motion on the SRS by projecting the paths on SH onto the fundamental domain SH/SΓ. We study closed orbits on the SRS first. A path $Z(t) = (z(t),\theta(t))$ on SH gives a closed loop on the SRS if it satisfies the condition that there exists an element $k \neq 1$ in SΓ and a time interval T such that

$$Z(t + T) = k(Z(t)). \tag{74}$$

Since k is characterized by the two fixed points, (u,μ) and (v,ν), the sign factor X and the norm function N (see Sec. II B),[23] the above condition gives a necessary condition,

$$\frac{z(t+T) - u - \theta(t+T)\mu}{z(t+T) - v - \theta(t+T)\nu} = N\frac{z(t) - u - \theta(t)\mu}{z(t) - v - \theta(t)\nu},$$
$$\frac{\theta(t+T) + [(\nu-\mu)/(u-v)]z(t+T) + (v\mu - u\nu)/(u-v)}{z(t+T) - v - \theta(t+T)\nu} = \chi N^{1/2}\frac{\theta(t) + [(\nu-\mu)/(u-v)]z(t) + (v\mu - u\nu)/(u-v)}{z(t) - v - \theta(t)\nu}. \tag{75}$$

We find that the classical motions $Z^{(\mathrm{II})}(t)$ (69) and $Z^{(\mathrm{III})}(t)$ (70) do not satisfy the above condition (75) and only the motions $Z^{(\mathrm{I})}(t)$ (68) with the parameters having values,[25,25]

$$c_1 = (v-u)/2, \quad c_2 = (u+v)/2, \quad \xi_2 = \nu/2,$$
$$\xi_4 = \mu/2, \quad \xi_1 = \xi_3 = 0, \tag{76}$$

satisfy the original condition (74) and the time interval T is

$$T = \log N/\omega. \tag{77}$$

Note that it is better to examine the limit $t \to \pm\infty$ first in order to get the condition (76). The path $Z^{(\mathrm{I})}(t)$ with (76), which we denote $Z_k(t)$ associated with the element k, is the geodesic curve connecting the two fixed points of the element $k \neq 1$ in $S\Gamma$;

$$Z^{(\mathrm{I})}(t \to +\infty) \to (v,\nu),$$
$$Z^{(\mathrm{I})}(t \to -\infty) \to (u,\mu), \quad \omega > 0. \tag{78}$$

A segment $[Z_k(t), Z_k(t+T)]$ of the geodesic curve becomes a closed loop on the SRS and the length of the loop $l(k)$ is given by

$$l(k) \equiv d(Z_k(t), Z_k(t+T)) = d(Z_k(t), k(Z_k(t))) = \log N, \tag{79}$$

which in fact depends only on the element $k \in S\Gamma$. Equation (79) yields

$$d(Z_k(t), Z_k(t+nT)) = d(Z_k(t), k^n(Z_k(t))) = |n| \log N, \tag{80}$$

or,

$$l(k^n) = |n| l(k). \tag{81}$$

The geodesic segment $[Z_k(t), Z_k(t+nT)]$ becomes a closed loop lying $|n|$-fold exactly on the closed loop coming from the segment $[Z_k(t), Z_k(t+T)]$. So $[Z_k(t), Z_k(t+nT)]$, and $[Z_k(t), Z_k(t+T)]$ determine the same periodic orbit, and we conclude that two elements k^m and $k^n \neq 1$ (m,n: integers) in $S\Gamma$ are associated with the same periodic orbit on the SRS. Furthermore, due to SPL(2,$\mathbb{R}$) invariance of $d(q_1,q_2)$,

$$d(\tilde{q}_1, \tilde{q}_2) = d(q_1, q_2), \tag{82}$$

we get

$$l(k) = d(gZ_k(t), gZ_k(t+T))$$
$$= d(gZ_k(t), gkg^{-1}(g\,Z_k(t))), \quad g \in S\Gamma. \tag{83}$$

This implies that $gZ_k(t)$ is the geodesic curve connecting the two fixed points of the element $gkg^{-1} \in S\Gamma$. Since $gZ_k(t)$ and $Z_k(t)$ become the same trajectory on the SRS, we conclude that every geodesic curve connecting the fixed points of each elements of a conjugacy class $\{k\}$ in $S\Gamma$ becomes the same orbit on the SRS. Thus we find that each pair of inconjugate primitive elements (p, p^{-1}) is associated with a periodic orbit on the SRS with its length given by

$$l(p) = \log N_p = \log N_{p^{-1}}, \tag{84}$$

where N_p is the norm function associated with p. Conversely, any periodic orbit can be lifted to a geodesic segment $[Z(t), k(Z(t))]$ on SH with some element $k \neq 1$ in $S\Gamma$. Since there exists a unique geodesic curve connecting the two points $Z(t)$ and $k(Z(t)) = Z(t+T)$, the geodesic curve is in fact a solution $Z^{(\mathrm{I})}(t)$ connecting the two fixed points of k. Then we conclude that there exists a one-to-one correspondence between periodic orbits on the SRS and pairs of inconjugate primitive elements (p, p^{-1}). Any geodesic curve $Z^{(\mathrm{I})}(t)$ not connecting two fixed points of any element in $S\Gamma$ becomes a nonperiodic orbit on the SRS and such geodesic curves are dense on SH. Hence, the classical motion on the SRS is chaotic, which we will discuss below.

The Lagrangian (41) is SPL(2,$\mathbb{R}$)-invariant, however, after projecting out onto the SRS, we find that the symmetry generators on SH no longer become those on the SRS and only two Grassmann even quantities are conserved, which are the Hamiltonian H and a nilpotent quantities $H^{(2)}$ essentially corresponding to $E^{\theta} E^{\bar{\theta}}$.[26] The fact that there are two kinds of conversed quantities has been already presented in constructing the Lagrangain which consists of two SPL(2,$\mathbb{R}$)-invariant pieces. However, the dimension of the hypersurface determined by $H = E$ and $H^{(2)} = E^{(2)}$($E, E^{(2)}$: constants) in the total superspace becomes less by one bosonic degree than that of total space according to the (super) implicit function theorem.[27]

We will study the Anosov property[28] that describes the behavior at large times of initially neighboring trajectories and is suitable to study strongly chaotic systems.[29] Let us take two geodesic curves $Z^{(\mathrm{I})}(t)$ ($\omega > 0, t_0 = 0$) with the conditions (76) and other ones, respectively,

$$c_1 = (v + \delta v - u)/2, \quad c_2 = (u + v + \delta v)/2,$$
$$\xi_2 = (\nu + \delta\nu)/2, \tag{85}$$
$$\xi_4 = \mu/2, \quad \xi_1 = \xi_3 = 0.$$

These two trajectories start from the same point (u,μ) at $t \to -\infty$, however, arrive at slightly different points (v,ν) and $(v+\delta v, \nu + \delta\nu)$ when $t \to \infty$. At $t=0$ the value of separation is

$$d_{t=0} \sim \delta v + [(\mu + \nu)/2]\delta\mu. \tag{86}$$

However, as $t \to \infty$ the trajectories separate exponentially,

$$d_{t \to \infty} \sim (\delta v + \nu\,\delta\nu) e^{\omega t}. \tag{87}$$

The velocity ω is the Liapunov exponent. This implies that trajectories are unstable, which is characteristic of classical chaos. Here the Liapunov exponent is in fact the Kolmogolov–Sinai entropy[30] h which, roughly speaking, measures unpredictability of the motions. This number comes out in the asymptotic formula for the counting function of primitive orbits of period (or loop length) $T(p) \leqslant T$,

$$\#\{p, T(p) \leqslant T\} \sim e^{hT}/hT, \quad T \to +\infty, \tag{88}$$

which indicates the exponential proliferation of the periodic orbits. This formula holds for Anosov systems in general.[31–33] We have seen the correspondence between periodic orbits and conjugacy classes of $S\Gamma$,

$$\#\{\text{periodic orbits}\} \simeq \#\{\text{conjugacy classes}, Q\}. \tag{89}$$

Then, the asymptotic formula (88) will be calculated through the properties of the super zeta function (see Sec. V) along the arguments in Refs. 34,35. It has not been examined strictly due to the existence of Grassmann odd numbers, however, a naive consideration supports the fact.

IV. QUANTUM MECHANICS

In this section we develop quantization. We first give the quantum Hamiltonian. Our Langrangian (41) is nonlinear in a sense that g_{AB} are functions of supercoordinates. Omote and Sato[36] have developed a procedure to construct the Hamiltonian for a system with a (purely bosonic) nonlinear Lagrangian of a form $L_B = \frac{1}{2}g_{ij}(q)\dot{q}^i\dot{q}^j$ symmetry as a guiding principle (see also Ref. 37). We can follow their arguments paying attention to sign factors. We find the quantum Hamiltonian,

$$H_Q = [(-)^A/2m]g^{-1/4}p_A g^{1/2}g^{AB}p_B g^{-1/4}, \tag{90}$$

where

$$g \equiv |\mathrm{sdet}\, g_{AB}| = (4a^2Y^2)^{-1}, \tag{91}$$

with the canonical commutation relations,

$$[p_A, q^B]_\pm = -i\hbar\delta_A^B. \tag{92}$$

The Hamiltonian H_Q is also SPL (2,$\mathbb{R}$)-invariant.

The scalar product for superfunctions on SH is

$$\langle\Psi_1|\Psi_2\rangle \equiv \int d^4q\, g^{1/2}(q)\langle\Psi_1|q\rangle\langle q|\Psi_2\rangle = \int dV\,\overline{\Psi}_1\Psi_2, \tag{93}$$

where $|q\rangle$ is an eigenfunction of the coordinate operator q^A with an eigenvalue q^A, which satisfies

$$I = \int d^4q\, g^{1/2}(q)|q\rangle\langle q|,$$
$$\langle q|q\rangle = [g(q)g(q')]^{-1/4}\delta^4(q-q'), \tag{94}$$

with the volume element dV given in (40). In the q-representation, the coordinates q^A and momenta p_A are given by

$$q^A = q^A,$$
$$p_A = -i\hbar g^{-1/4}\frac{\partial}{\partial q^A}g^{1/4} \equiv -i\hbar g^{-1/4}\,\partial_A\, g^{1/4}, \tag{95}$$

so that H_Q is a super Laplace–Beltrami (SLB) operator,

$$H_Q = -(\hbar^2/2m)(-)^A g^{-1/2}\partial_A(g^{1/2}g^{AB}\partial_B)$$
$$= -(\hbar^2/2m)[(2YD\overline{D})^2 + (1-a)/a(2YD\overline{D})]$$
$$\equiv (\hbar^2/2m)\Delta_{\mathrm{SLB}}, \tag{96}$$

where D is given in (6) and

$$\overline{D} \equiv -\partial_{\bar\theta} + \bar\theta\,\partial_{\bar z}. \tag{97}$$

We will study the spectral properties of H_Q on SH. In order to do so, we examine the eigenvalue problem of the operator $\Box_0$,

$$\Box_0 \equiv 2YD\overline{D}. \tag{98}$$

First, we consider the Grassmann even eigenfunction e_Λ, with eigenvalue λ,

$$\Box_0 e_\Lambda = \lambda e_\Lambda, \tag{99}$$

where e_Λ may be expanded with respect to the Grassmann odd coordinates θ and $\bar\theta$ as,

$$e_\Lambda = A_\Lambda + \theta\bar\theta B_\Lambda, \tag{100}$$

where A_Λ and B_Λ are function of Grassmann even coordinates $z = x + iy$ and $\bar z$. Equations (99) and (100) yield

$$B_\Lambda = (\lambda/2y)A_\Lambda,$$
$$\{y^2(\partial_x^2 + \partial_y^2) - \lambda(\lambda-1)\}A_\Lambda = 0. \tag{101}$$

We find that A_Λ is solved as,[38]

$$A_\Lambda = C_{\lambda,k}e^{ikx}\sqrt{y}K_{\lambda-1/2}(|k|y), \tag{102}$$

where $C_{\lambda,k}$ is a normalization constant determined below and $K_{\lambda-1/2}(|k|y)$ is a modified Bessel function which damps exponentially as $|y|\to\infty$,

$$\{\partial_z^2 + (1/z)\,\partial_z - (1+\nu^2/z^2)\}K_\nu(z) = 0. \tag{103}$$

We shall examine the normalization condition. Since SH is noncompact and the spectrum, which is parametrized by λ and k, is continuous, the normalization condition should be

$$\langle e_\Lambda|e_{\Lambda'}\rangle \propto \delta(\lambda-\lambda'). \tag{104}$$

The above condition determines the region in which λ runs. Details are given in Appendix D. We find that

$$\lambda = \tfrac{1}{2} + ip, \quad p\in(-\infty, +\infty). \tag{105}$$

Thus the Grassmann even eigenfunctions are given by

$$e_{p,k}(Z) = \left(\frac{2ia\sinh\pi p}{\pi^3}\right)^{1/2}\left(1 + \frac{1+2ip}{4y}\theta\bar\theta\right)$$
$$\times e^{ikx}\sqrt{y}K_{ip}(|k|y), \tag{106}$$

which satisfy,

$$\Box_0 e_{p,k} = (ip+\tfrac{1}{2})e_{p,k}, \tag{107}$$
$$\langle e_{q,l}|e_{p,k}\rangle = \delta(p+q)\delta(k-l), \tag{108}$$

where a point $(p,k) = (0,0)$ is understood to be excluded. Hence $e_{p,k}(Z)$ is an eigenfunction of H_Q with an eigenvalue $E^B_{p,k}$,

$$E^B_{p,k} = \frac{\hbar^2}{2m}\left\{\left(\frac{1-a}{2a}\right)^2 + \left(p - \frac{i}{2a}\right)^2\right\} \equiv \frac{\hbar^2}{2m}\gamma^B(p). \tag{109}$$

Although H_Q is a hermite operator, the eigenvalue is complex. This is because the space of eigenstates contains isovectors[15] as is seen in (108).

Next, we will proceed to the Grassmann odd eigenfunction ψ_Λ which may be expanded as,

$$\psi_\Lambda = (1/\sqrt{y})(\theta\rho_\Lambda + \bar\theta\varphi_\Lambda). \tag{110}$$

The equation,

$$\Box_0\psi_\Lambda = \lambda\psi_\Lambda, \tag{111}$$

yields

$$\varphi_\Lambda = -(2y\sqrt{y}/\lambda)\partial_{\bar z}(\rho_\Lambda/\sqrt{y}),$$
$$\{y^2(\partial_x^2 + \partial_y^2) - iy\,\partial_x - (\lambda^2 - \tfrac{1}{4})\}\rho_\Lambda = 0. \tag{112}$$

These differential equations can be solved as (see Appendix E),

$$\rho_\Lambda = C_{\lambda,k}e^{ikx}W_{(\sigma_k/2),\lambda}(2|k|y),$$
$$\varphi_\Lambda = iC_{\lambda,k}\lambda^{\sigma_k}e^{ikx}W_{-(\sigma_k/2),\lambda}(2|k|y), \tag{113}$$

where

$$\sigma_k \equiv \text{sign}(k), \quad k \neq 0, \tag{114}$$

and $W_{\kappa,\mu}$ is a Whittaker function which satisfies

$$\left\{\partial_z^2 + \left(-\frac{1}{4} + \frac{\kappa}{z} - \frac{\mu^2 - 1/4}{z^2}\right)\right\} W_{\kappa,\mu}(z) = 0. \tag{115}$$

The normalization condition determines the region where the parameter λ runs as in the Grassmann even case. We find (see Appendix E) surprisingly that in general the region does not coincide with the one for the Grassmann even eigenfunctions (105),

$$\lambda = c + ip, \tag{116}$$

where

$$c\text{: real constant}, \quad |c| \leqslant \tfrac{1}{2}, \tag{117}$$

$$p \in (-\infty, +\infty),$$

and the eigenfunctions are

$$\psi^c_{p,k}(Z) = \left(\frac{a\cos[\pi(c+ip)]}{2\pi^2 k(c+ip)^{\sigma_k - 1}}\right)^{1/2} \frac{1}{\sqrt{y}} e^{ikx} \times \{\theta W_{\sigma_k/2, c+ip}(2|k|y) + i(c+ip)^{\sigma_k}\bar\theta W_{-\sigma_k/2, c+ip}(2|k|y)\}, \tag{118}$$

which satisfy

$$\Box_0 \psi^c_{p,k} = (c+ip)\psi^c_{p,k},$$

$$\langle \psi^c_{q,l} | \psi^c_{p,k} \rangle = \delta(k-l)\delta(p+q). \tag{119}$$

The eigenvalues of H_Q are

$$E^F_{p,k,c} = \frac{\hbar^2}{2m}\left\{\left(\frac{1-a}{2a}\right)^2 + \left(p - ic - i\frac{1-a}{2a}\right)^2\right\} \equiv \frac{\hbar^2}{2m}\gamma^F_c(p). \tag{120}$$

Notice that except when $c = \frac{1}{2}$ or $(a-2)/2a$ with $a \geqslant 1$, the energy spectra of the Grassmann even states and the odd ones do not coincide with each other,

$$\{E^B_{p,k}\} \neq \{E^F_{q,l,c}\}, \quad c \neq \tfrac{1}{2} \text{ and } (a-2)/2a \quad (a \geqslant 1). \tag{121}$$

A set of eigenfunctions for each c,

$$\{e_{p,k}, \psi^c_{p,k}\}, \tag{122}$$

do satisfy the completeness relation (see Appendix F),

$$\int_{-\infty}^{\infty} dp\, dk\, [e_{p,k}(q_2)\bar e_{-p,k}(q_1) + \psi^c_{p,k}(q_2)\bar\psi^c_{-p,k}(q_1)] = 2a(Y_{(1)}Y_{(2)})^{1/2}\delta(x_1 - x_2)\delta(y_1 - y_2) \times (\bar\theta_1 - \bar\theta_2)(\theta_1 - \theta_2) = [g(q_1)g(q_2)]^{-1/4}\delta(q_1 - q_2). \tag{123}$$

For each c, we have a Hilbert space for the Grassmann odd states $\mathcal{H}^F_c$, and hence the total Hilbert space is

$$\mathcal{H}_c = \mathcal{H}^B \otimes \mathcal{H}^F_c, \tag{124}$$

where $\mathcal{H}^B$ is the Hilbert space for the Grassmann even states.

Since we have gotten the eigenfunctions on SH we can get those on the SRS by the Poincaré series,

$$\Phi_{\text{SRS}}(q) = \sum_{g \in \Gamma} \Phi_{\text{SH}}(g(q)), \tag{125}$$

however, because the summation is complicated it is difficult to see the spectrum on the SRS from those eigenfunctions.

We will construct the kernel function on SH. The kernel function is given by,

$$K(q_1,q_2;t) \equiv \langle q_2 | e^{-(it/\hbar)H_Q} | q_1 \rangle = \int_{-\infty}^{\infty} dp\, dk \{e^{-(it/\hbar)E^B_{p,k}} \langle q_2 | e_{p,k}\rangle \langle e_{-p,k} | q_1 \rangle + e^{-(it/h)E^F_{p,k,c}} \langle q_2 | \psi^c_{p,k}\rangle \langle \psi^c_{-p,k} | q_1\rangle\}$$
$$= \int_{-\infty}^{\infty} dp\, dk \left\{e^{-\tau\gamma^B(p)} e_{p,k}(q_2)\bar e_{-p,k}(q_1) + e^{-\tau\gamma^F_c(p)}\psi^c_{p,k}(q_2)\bar\psi^c_{-p,k}(q_1)\right\} \equiv K(q_1,q_2|\tau), \tag{126}$$

where

$$\tau \equiv (i\hbar/2m)t. \tag{127}$$

Plugging (106) and (118) into (126), we finally get (see Appendix G)

$$K(q_1,q_2|\tau) = K^{(0)}(l;\tau) + r(q_1,q_2)K^{(1)}(l;\tau), \tag{128}$$

where $l = l(q_1,q_2)$ is given in (65) and,

$$K^{(0)}(l;\tau) = \frac{-2a}{\pi\sqrt{2\pi\tau}} e^{-((1-a)/2a)^2\tau} \int_l^{\infty} db \frac{e^{-b^2/4\tau}\sinh(b/2a)}{(\cosh b - \cosh l)^{1/2}}, \tag{129}$$

$$K^{(1)}(l;\tau) = \frac{-2a}{\pi\sqrt{2\pi\tau}} e^{-((1-a)/2a)^2\tau} \int_l^{\infty} db \frac{1}{(\cosh b - \cosh l)^{1/2}} \times \left[(\cosh l - 1)\frac{d}{db}\left(e^{-b^2/4\tau}\frac{\sinh(b/2a)}{\sinh b}\right) + e^{-b^2/4\tau}\left(\frac{b}{2\tau}\cosh\frac{b}{2a} + \frac{a-1}{2a}\sinh\frac{b}{2a}\right)\right]. \tag{130}$$

When $a = 1$, the above equations coincide with the corresponding ones of Ref. 14. So the time development for a wave function on SH is given by

$$\Psi(q,t) = \int_{SH} dV(q') K(q,q';t-t')\Psi(q',t'). \quad (131)$$

As for a wave function on the SRS or F, a fundamental domain of SΓ, we should have, similarly,

$$\Psi_{SRS}(q,t) = \int_{SRS} dV(q') K_{SRS}(q,q';t-t')\Psi_{SRS}(q',t'). \quad (132)$$

This implies that a kernel K on SH, or SH$\times$SH, induces a kernel K_{SRS} on the SRS, or SH/S$\Gamma\times$SH/SΓ,

$$K_{SRS}(q_1,q_2|\tau) = \sum_{g\in S\Gamma} K(q_1,g(q_2)|\tau). \quad (133)$$

V. TRACE FORMULA AND ZETA FUNCTIONS

In the preceding section, we have eventually solved the quantum mechanics on SH and obtained the heat kernel on SH, which yields that on the SRS. Here we concentrate on the quantum energy spectrum on the SRS. In the case of a particle on a Riemann surface of genus $h>2$, it is related to the length spectrum through the Selberg trace formula.[2] We may expect that a similar relation will exist in our model.

First, we present a formula for the supertrace of a function $G_{SRS}(q_1,q_2)$ on SH/S$\Gamma\times$SH/SΓ which is made out of a SPL(2,**R**)-invariant function $G(q_1,q_2)$ on SH$\times$SH;

$$G_{SRS}(q_1,q_2) = \sum_{g\in S\Gamma} G(q_1,g(q_2)), \quad (134)$$

with

$$G(q_1,q_2) = \Phi(l(q_1,q_2)) + r(q_1,q_2)\Psi(l(q_1,q_2)), \quad (135)$$

where Φ and Ψ are some functions and l and r are given in (65) and (62), respectively. We find that the supertrace of G_{SRS} defined by,

$$\mathrm{str}\, G_{SRS} = \int_F dV(q) G_{SRS}(q,q) = \sum_{g\in S\Gamma}\int_F dV(q) G(q,g(q)), \quad (136)$$

where F is a fundamental domain of SΓ, is calculated as

$$\mathrm{str}\, G_{SRS} = \mathrm{Area(SRS)}\ \Phi(0) + \sum_{\substack{\text{inconjugate}\\ \text{primitive } p}} \sum_{n=1}^{\infty}\int_{F_p} dV\, G(q,g(q)), \quad (137)$$

where use has been made of (19) and F_p is a fundamental domain for the centralizer $Z(p)$,

$$F_p = \bigcup_{g\in S\Gamma} g^{-1}F. \quad (138)$$

We can assume that p is a magnification with a matrix $A_p = \mathrm{diag}(\chi(p)N_p^{1/2}, \chi(p)N_p^{-1/2}, 1)$ [see (14)] and we can choose a convenient domain for F_p,[21]

$$\int_{F_p} dV \Rightarrow \int_1^{N_p} dy \int_{-\infty}^{\infty} dx \int d\theta\, d\bar\theta (2ay + a\theta\bar\theta)^{-1}. \quad (139)$$

Then, we finally get

$$\mathrm{str}\, G_{SRS} = \frac{\pi(h-1)}{a}\Phi(0) - \sum_{\substack{\text{inconjugate}\\ \text{primitive } p}} \sum_{n=1}^{\infty} \frac{l(p)}{2a\sqrt{\cosh l(p^n) - 1}} \left\{\left(1 - \chi(p^n)\cosh\frac{l(p^n)}{2}\right) \times \int_{l(p^n)}^{\infty} ds \frac{\sinh s}{(\cosh b - \cosh l(p^n))^{1/2}} \Psi(s) + (1-\cosh l(p^n)) \int_{l(p^n)}^{\infty} ds \frac{1}{(\cosh b - \cosh l(p^n))^{1/2}} \frac{d\Phi(s)}{ds}\right\}, \quad (140)$$

here h is the genus of the SRS and

$$l(p^n) = |n| l(p) = |n| \log N_p. \quad (141)$$

Now we apply the above formula to the heat kernel on the SRS (133) which can be written as

$$K_{SRS}(q_1,q_2|\tau) = \sum_{g\in S\Gamma} \langle q_1 | e^{-\tau\Delta_{SLB}} | g(q_2)\rangle. \quad (142)$$

The spectrum of the operator Δ_{SLB} on SH, $\{\gamma^B(p)\}$ and $\{\gamma^F(p)\}$,[39] will become discrete on the SRS and let $\{\gamma_n^B\}$ and $\{\gamma_n^F\}$ $(n=0,1,2,\ldots)$ be the eigenvalues for Grassman even and odd functions, respectively. We then get

$$\mathrm{str}\, K_{SRS} = \mathrm{str}(e^{-\tau\Delta_{SLB}}) = \sum_{n=0}^{\infty}(e^{-\tau\gamma_n^B} - e^{-\tau\gamma_n^F}). \quad (143)$$

Plugging $K^{(0)}$ (129) and $K^{(1)}$ (130), respectively, into Φ and Ψ in (140) and integrating with respect to s, we finally obtain a superanalog of the Selberg trace formula,

$$\sum_{n=0}^{\infty}(e^{-\tau\gamma_n^B} - e^{-\tau\gamma_n^F}) = A(\tau) + \Theta(\tau), \quad (144)$$

where

$$A(\tau) = \frac{1}{\sqrt{\pi\tau}}(1-h)e^{-((1-a)/2a)^2\tau} \times \int_0^{\infty} db\, e^{-b^2/4\tau} \frac{\sinh(b/2a)}{\sinh(b/2)}, \quad (145)$$

$$\Theta(\tau) = \frac{1}{4\sqrt{\pi\tau}} \sum_{\substack{\text{inconjugate}\\ \text{primitive } p}} \sum_{n=1}^{\infty} \mathrm{str}(K(p^n|a)) \times \frac{l(p)}{\sinh(l(p^n)/2)} \exp[-l^2(p^n)/4\tau - ((1-a)/2a)^2\tau], \quad (146)$$

$$K(p|a) = \mathrm{diag}(e^{l(p)/2a}, e^{-l(p)/2a}, \chi(p)e^{[(1-a)/2a]l(p)}, \chi(p)e^{-[(1-a)/2a]l(p)}). \quad (147)$$

Here, $A(\tau)$ is the contribution of "trivial motion" on the SRS (zero length term) and for $\tau \to 0$ it can be expanded into a positive power series,

$$A(\tau) \sim -\frac{\mathrm{Area(SRS)}}{\pi}(b_0 + b_1\tau + b_2\tau^2 + \cdots), \quad (148)$$

with

$$b_0 = 1,$$
$$b_1 = \frac{(a-1)(1-2a)}{6a^2},$$
$$b_2 = \frac{(a-1)(2a-1)(2a^2-2a+1)}{60a^4},$$
$$b_3 = \frac{(a-1)(1-2a)(16a^4-18a^3+21a^2-12a+3)}{2520a^6},$$
$$\vdots \quad (149)$$

This series approximates "$\mathrm{str}(e^{-\tau\Delta_{\mathrm{SLB}}})$" up to an exponentially small error and each coefficient b_n is written by the Riemann curvature tensors (see Appendix H);

$$b_1 = R/6$$
$$b_2 = (1/360)(5R^2 - 2R_{AB}R^{BA} + 2R_{ABCD}R^{DCBA}),$$
$$b_3 = (1/1128)(620R^3 + 16RR_{AB}R^{BA} + 100RR_{ABCD}R^{DCBA} + 460R_{AB}R^{BA}_{\ \ CD}R^{DC} - 285R_{ABCD}R^{CD}_{\ \ EF}R^{FEBA}),$$
$$\vdots \quad (150)$$

On the other hand, $\Theta(\tau)$ is the contribution from the periodic motions on the SRS and consistent with the semiclassical approximation.[9]

Next we discuss super zeta functions associated with our model. Let us introduce two kinds of zeta functions $Z(s|a)$[10] and $Z_\Delta(s|a)$ with one parameter a corresponding to the Selberg zeta function and the Minakshisundaram–Pleijel zeta function[40] in the theory of classical Riemann surfaces. These functions are defined by,

$$Z(s|a) \equiv \prod_{\substack{\text{inconjugate}\\ \text{primitive } p}} \prod_{n=0}^{\infty} \mathrm{sdet}(1 - K(p|a)e^{-(s+n)l(p)}), \quad (151)$$

$$Z_\Delta(s|a) \equiv \prod_{n=1}^{\infty} \{(\gamma_n^B)^{-s} - (\gamma_n^F)^{-s}\}. \quad (152)$$

First, we show that the zeta functions are related to the trace formula;

I. $$\frac{d}{ds}\log Z(s|a) = (2s-1) \times \int_0^\infty dt \exp[-(s-1/2)^2 t + ((1-a)/2a)^2 t]\Theta(t), \quad (153)$$

II. $$\Gamma(s)Z_\Delta(s|a) = \int_0^\infty dt \{\mathrm{str}(e^{-t\Delta_{\mathrm{SLB}}}) - M\}t^{s-1},$$

where

$$M \equiv \lim_{t\to\infty} \mathrm{str}(e^{-t\Delta_{\mathrm{SLB}}}). \quad (154)$$

Proof: (I) By introducing matrix components $Z(s)$ and $\tilde{Z}(s)$,[21]

$$Z(s) = \prod_{\substack{\text{inconjugate}\\ \text{primitive } p}} \prod_{n=0}^{\infty}(1 - e^{-(s+n)l(p)}),$$
$$\tilde{Z}(s) = \prod_{\substack{\text{inconjugate}\\ \text{primitive } p}} \prod_{n=0}^{\infty}(1 - \chi(p)e^{-(s+n)l(p)}), \quad (155)$$

we can rewrite $Z(s|a)$ as

$$Z(s|a) = \frac{Z(s+1/2a)Z(s-1/2a)}{\tilde{Z}(s+(1-a)/2a)\tilde{Z}(s-(1-a)/2a)}. \quad (156)$$

We then find[4,34]

$$\frac{d}{ds}\log Z(s) = \frac{1}{2}\sum_{\substack{\text{inconjugate}\\ \text{primitive } p}} \sum_{n=0}^{\infty} l(p)\frac{e^{-(s-1/2)l(p^n)}}{\sinh[l(p)/2]},$$
$$\frac{d}{ds}\log \tilde{Z}(s) = \frac{1}{2}\sum_{\substack{\text{inconjugate}\\ \text{primitive } p}} \sum_{n=0}^{\infty} l(p)\chi(p^n)\frac{e^{-(s-1/2)l(p^n)}}{\sinh[l(p)/2]}. \quad (157)$$

Combining (156) and (157) we can deduce the formula in question;

$$\frac{d}{ds}\log Z(s|a) = \frac{1}{2}\sum_{\substack{\text{inconjugate}\\ \text{primitive } p}} \sum_{n=0}^{\infty} \frac{l(p)}{\sinh[l(p)/2]}\mathrm{str}(K(p^n|a))e^{-(s-1/2)l(p^n)}$$
$$= \left(s-\frac{1}{2}\right)\sum_{\substack{\text{inconjugate}\\ \text{primitive } p}} \sum_{n=0}^{\infty} \frac{1}{\sqrt{4\pi}}\mathrm{str}(K(p^n|a))\frac{l(p)}{\sinh[l(p)/2]}\int_0^\infty dt\frac{1}{\sqrt{t}}\exp[-l^2(p^n)/4t - (s-1/2)^2 t]$$
$$= (2s-1)\int_0^\infty dt \exp[-(s-1/2)^2 t + ((1-a)/2a)^2 t]\Theta(t). \quad (158)$$

Q.E.D.

(II) The proof follows immediately due to the Mellin transformation. Q.E.D.

Secondly, we point out that the zero-points and the poles of $Z(s|a)$ directly give the eigenvalues of $\Delta_{\rm SLB}$ (energy spectrum) on the SRS. More precisely, the zero points give the eigenvalues of the Grassmann even functions and the poles give those of the odd functions. Using the trace formula, we find

$$\frac{d}{ds}\log Z(s|a) = (2s-1)\int_0^\infty dt\exp[-(s-1/2)^2t+((1-a)/2a)^2t]\left\{\sum_{n=0}^{\infty}(e^{-t\gamma_n^B}-e^{-t\gamma_n^F})-A(t)\right\}$$

$$= (2s-1)\sum_{n=0}^{\infty}\left[\frac{1}{(s-\frac{1}{2})^2-((1-a)/2a)^2+\gamma_n^B}-\frac{1}{(s-\frac{1}{2})^2-((1-a)/2a))^2+\gamma_n^F}\right]$$

$$+2(h-1)\sum_{n=0}^{\infty}\left(\frac{1}{s+n-1/2a}-\frac{1}{s+n+1/2a}\right). \quad (159)$$

Note that the last term in (159) becomes a finite sum when $|a|^{-1}=m$ (m: positive integers),

$$2(h-1)\mathrm{sign}(a)\sum_{n=0}^{m-1}\frac{1}{s+n-m/2}. \quad (160)$$

This formula implies that $Z(s|a)$ has a meromorphic continuation onto the whole complex plane $\mathbb{C}$. The zero points (poles) of order 1 exist at

$$s=\tfrac{1}{2}\pm\sqrt{((1-a)/2a)^2-\gamma_n^{B(F)}}. \quad (161)$$

Other trivial zero points (ZP) and poles (P) exist respectively at,

I. $a^{-1}\neq m$ (m: positive integers);

ZP: $s=n+1/2a$,

P: $s=-n-1/2a \quad (n=0,1,2,...)$, (162)

II. $a^{-1}=m$;

ZP: $s=-n+m/2$,

P: none $(n=0,1,...,m-1)$, (163)

III. $a^{-1}=-m$;

ZP: none,

P: $s=-n-m/2$
$(n=0,1,...,m-1)$, (164)

where the order of each ZP and P is $2(h-1)$.

$Z_\Delta(s|a)$ may be continued analytically onto the whole complex plane $\mathbb{C}$. In fact, we have

$$\Gamma(s)Z_\Delta(s|a)=\int_0^1 dt\{\mathrm{str}(e^{-t\Delta_{\rm SLB}})-M\}t^{s-1}$$
$$+\int_1^\infty dt\{\mathrm{str}(e^{-t\Delta_{\rm SLB}})-M\}t^{s-1}. \quad (165)$$

The second integral converges for any $s\in\mathbb{C}$ since the integrand damps exponentially for $t\to\infty$. The first term can be expanded according to (148) and (149) for $t\to 0$ as,

$$\int_0^1 dt\{\mathrm{str}(e^{-t\Delta_{\rm SLB}})-M\}t^{s-1}$$
$$=-\frac{\mathrm{Area(SRS)}}{\pi}\int_0^1 dt\{b_0t^{s-1}+b_1t^s+b_3t^{s+1}+\cdots\}-\frac{M}{s}$$
$$=\frac{1-h}{a}\left[\frac{b_0}{s}+\frac{b_1}{s+1}+\frac{b_2}{s+2}+\cdots\right]-\frac{M}{s}. \quad (166)$$

Since $\Gamma(s)$ has a simple pole at every nonpositive integer $n\leqslant 0$, $Z_\Delta(s|a)$ is analytic.

Finally we derive a functional equation for $Z(s|a)$;

$$\frac{Z(s|a)}{Z(1-s|a)}=\left\{\frac{\sin\pi(s-1/2a)}{\sin\pi(s+1/2a)}\right\}^{2(h-1)}, \quad (167)$$

where h is the genus of the SRS. Note that when $|a|^{-1}=$ integer, the rhs becomes 1 and hence $Z(s|a)$ is symmetric under the interchange of $s\leftrightarrow 1-s$.

Proof: Let

$$F(s,a)\equiv 1/(2s-1)\frac{d}{ds}\log Z(s|a). \quad (168)$$

Due to (159) we have

$$F(s,a)-F(1-s,a)$$
$$=\frac{h-1}{s-\frac{1}{2}}\sum_{n=-\infty}^{\infty}\left(\frac{1}{s+n-1/2a}-\frac{1}{s+n+1/2a}\right)$$
$$=\frac{\pi(h-1)}{s-\frac{1}{2}}\{\cot\pi(s-1/2a)-\cot\pi(s+1/2a)\}, \quad (169)$$

which leads to

$$\frac{d}{ds}\log\frac{Z(s|a)}{Z(1-s|a)}$$
$$=2\pi(h-1)\{\cot\pi(s-1/2a)-\cot\pi(s+1/2a)\}. \quad (170)$$

Note that when $|a|^{-1}=m$, the rhs vanishes. The functional equation is obtained by integrating the above equation. Q.E.D.

VI. SUMMARY AND DISCUSSIONS

We have examined a free motion of a superparticle on a SRS of genus $h\geqslant 2$. We have taken SH, a superanalog of the complex upper half-plane, as a universal covering space of the SRS and hence the SRS is represented as a fundamental domain F of a super Fuchsian group SΓ. The classical periodic orbits on the SRS have a close relation with the SΓ. That is, we have seen that there exists a one-to-one correspondence between the periodic orbits and the pairs of inconjugate primitive elements and their inverse elements, with the lengths of the periodic orbits given by the norm function of the SΓ. Classical motions have chaotic aspects and the asymptotic form of the number of primitive periodic orbits is expected. Due to the above correspondence between periodic orbits and primitive elements, the asymptotic form yields that of the number of inconjugate primitive elements of the $S\Gamma$.

We have solved the quantum mechanics on the universal covering space SH. Both Grassmann even and odd eigenfunctions were given and the heat kernel was constructed by summing them up. The (Grassmann odd) eigenfunctions depend on a real parameter $c, |c| \leqslant \frac{1}{2}$, and hence there are infinite number of inequivalent total Hilbert spaces, parametrized by the parameter c, however, the kernel function is unique, or independent of the parameter c. Although the classical energy spectrum is real and the quantum Hamiltonian is a hermite operator, the quantum energy spectrum is complex. If E is an eigenvalue of the Hamiltonian, then $\bar{E}$ is also an eigenvalue. This is due to the existence of the isovectors.

As for the quantum mechanics on the SRS, we have given the exact kernel function. It leads to a superanalog of the Selberg trace formula (144). The supertrace of the "super" Laplace–Beltrami operators $\Delta_{\rm SLB}$ (96) consists of two parts, $A(t)$ (145) and $\Theta(t)$ (146). We see that $A(t)$ is the contribution of the zero-length "periodic orbits" and $\Theta(t)$ is of the nonzero classical periodic orbits. The latter is, in fact, given with the heat kernel in the semiclassical approximation up to some overall factor depending on the parameters a.[9,10] As for the former, it can also be calculated by the so-called heat kernel expansion (150) and the semiclassical kernel function gives the first term of the series. Furthermore, $\Theta(t)$ takes the same form as the one in the trace formula of a super Laplace–Beltrami operator with spin $(1-a)/a$,[21,15]

$$\Delta^{(m)}_{\rm SLB} = (2Y^{m/2+1}DY^{-m}\bar{D}Y^{m/2})^2 - [(1-a)/2a]^2. \tag{171}$$

This is not the case for $A(t)$.

The spectrum on the SRS is complicated, however, it can be investigated by the Selberg super zeta function (151). The zero points and poles of the super zeta function gives the spectra for the Grassmann even and odd states, respectively. Although the energy spectrum on SH depends on the parameter c and is different from each other for every value of c, we may expect that the spectrum on the SRS is independent of the parameter c and hence unique because the super zeta function is independent of c. We see that when $a \to \infty$, $\gamma^B(p)$ in (109) coincides with the eigenvalues of the Laplace–Beltrami operator on H in the theory of Riemann surfaces. Hence we may expect that almost all the nontrivial zero-points and poles of the super zeta function will exist on the line of $\mathrm{Re}\, s = \frac{1}{2}$ when $a \to \infty$, however, it may not be the case for general values of a. The super zeta function itself may be complicated, however, it satisfies a simple functional relation.

The spectrum of $\Delta_{\rm SLB}$ on the SRS is characterized by the length spectrum of the classical periodic orbits. In the theory of Riemann surfaces it was shown that some lengths of $6h-6$ can be chosen as coordinates of the Teichmüller space and the measure of the Teichmüller space was given in terms of the lengths.[41] In the case of SRS's similar situation may be expected and calculation of the differential of the lengths along the super Beltrami differentials was done.[42] Constructing a measure of the super Teichmüller space is important for superstring.

APPENDIX A

We give some geometrical quantities associated with our metric g_{AB} (39).

(a) Inverse metric g^{AB}:

$$(g^{AB}) = Y\begin{pmatrix} 0 & 2Y+\frac{1-2a}{a}\theta\bar\theta & 0 & \frac{a-1}{a}\theta+i\bar\theta \\ 2Y+\frac{1-2a}{a}\theta\bar\theta & 0 & \frac{a-1}{a}\bar\theta-i\theta & 0 \\ 0 & \frac{a-1}{a}\bar\theta-i\theta & 0 & \frac{1}{a} \\ \frac{a-1}{a}\theta+i\bar\theta & 0 & -\frac{1}{a} & 0 \end{pmatrix}. \tag{A1}$$

This satisfies

$$g_{AC}g^{CB} = g^{BC}g_{CA} = \delta^B_A. \tag{A2}$$

(b) Cristoffel symbol Γ^A_{BC}:

$$\Gamma^A_{BC} \equiv \tfrac{1}{2} g^{AD}[(-)^{B(1+D)}\partial_B g_{DC} + (-)^{C(1+B+D)}\partial_C g_{DB} - (-)^B \partial_D g_{BC}], \tag{A3}$$

$$(\Gamma^z_{AB}) = \begin{pmatrix} \frac{i}{Y} & \frac{(1-a)i\theta\bar\theta}{4aY^2} & \frac{\bar\theta}{2Y} & \frac{(a-1)\theta}{2aY} \\ \frac{(1-a)i\theta\bar\theta}{4aY^2} & 0 & 0 & 0 \\ \frac{\bar\theta}{2Y} & 0 & 0 & 0 \\ \frac{(a-1)\theta}{2aY} & 0 & 0 & 0 \end{pmatrix}, \tag{A4}$$

$$
(\Gamma^{\bar{z}}_{AB}) = \begin{pmatrix} 0 & \frac{(a-1)i\theta\bar{\theta}}{4aY^2} & 0 & 0 \\ \frac{(a-1)i\theta\bar{\theta}}{4aY^2} & -\frac{i}{Y} & \frac{(1-a)\bar{\theta}}{2aY} & -\frac{\theta}{2Y} \\ 0 & \frac{(1-a)\bar{\theta}}{2aY} & 0 & 0 \\ 0 & -\frac{\theta}{2Y} & 0 & 0 \end{pmatrix}, \tag{A5}
$$

$$
(\Gamma^{\theta}_{AB}) = \begin{pmatrix} 0 & \frac{(a-1)(\theta+i\bar{\theta})}{4aY^2} & \frac{i}{2Y} & \frac{(1-a)(2Y+\theta\bar{\theta})}{4aY^2} \\ \frac{(a-1)(\theta+i\bar{\theta})}{4aY^2} & 0 & \frac{(1-a)i\theta\bar{\theta}}{4aY^2} & 0 \\ \frac{i}{2Y} & \frac{(1-a)i\theta\bar{\theta}}{4aY^2} & 0 & \frac{(a-1)\theta}{2aY} \\ \frac{(1-a)(2Y+\theta\bar{\theta})}{4aY^2} & 0 & \frac{(1-a)\theta}{2aY} & 0 \end{pmatrix}, \tag{A6}
$$

$$
(\Gamma^{\bar{\theta}}_{AB}) = \begin{pmatrix} 0 & \frac{(1-a)(i\theta-\bar{\theta})}{4aY^2} & 0 & \frac{(a-1)i\theta\bar{\theta}}{4aY^2} \\ \frac{(1-a)(i\theta-\bar{\theta})}{4aY^2} & 0 & \frac{(1-a)(2Y+\theta\bar{\theta})}{4aY^2} & -\frac{i}{2Y} \\ 0 & \frac{(1-a)(2Y+\theta\bar{\theta})}{4aY^2} & 0 & \frac{(a-1)\bar{\theta}}{2aY} \\ \frac{(a-1)i\theta\bar{\theta}}{4aY^2} & -\frac{i}{2Y} & \frac{(1-a)\bar{\theta}}{2aY} & 0 \end{pmatrix}. \tag{A7}
$$

The following relation is useful,

$$
\Gamma_A \equiv (-)^{A+B}\Gamma^B_{BA} = \tfrac{1}{2}\partial_A(\log g) = -Y^{-1}\partial_A Y. \tag{A8}
$$

(c) Riemann tensor:

$$
R^{A}{}_{BCD} \equiv (-)^{C(A+B+1)}\partial_C\Gamma^A_{BD} - (-)^{D(A+B+C+1)}\partial_D\Gamma^A_{BC} + (-)^{C(B+E)}\Gamma^A_{EC}\Gamma^E_{BD} - (-)^{D(B+C+E)}\Gamma^A_{ED}\Gamma^E_{BC}. \tag{A9}
$$

(d) Ricci tensor:

$$
R_{AB} \equiv (-)^{C(A+1)}R_{ACB}{}^{C}
$$

$$
= (-)^C\partial_C\Gamma^C_{AB} - (-)^{AB+A+B}\partial_B\Gamma_A + (-)^C\Gamma_C\Gamma^C_{AB} - (-)^{(A+1)E}\Gamma^E_{AC}\Gamma^C_{EB}, \tag{A10}
$$

$$
(R_{AB}) = \begin{pmatrix} 0 & \frac{1-2a^2}{4a^2Y^2} & 0 & \frac{(3a-2)ai\theta-(1-a)^2\bar{\theta}}{4a^2Y^2} \\ \frac{1-2a^2}{4a^2Y^2} & 0 & \frac{(1-a)^2\theta+(3a-2)ai\bar{\theta}}{4a^2Y^2} & 0 \\ 0 & \frac{(1-a)^2\theta+(3a-2)ai\bar{\theta}}{4a^2Y^2} & 0 & \frac{2(2-3a)Y-(1-2a)\theta\bar{\theta}}{4aY^2} \\ \frac{(3a-2)ai\theta-(1-a)^2\bar{\theta}}{4a^2Y^2} & 0 & \frac{(1-2a)\theta\bar{\theta}-2(2-3a)Y}{4aY^2} & 0 \end{pmatrix}. \tag{A11}
$$

(e) Scalar curvature:

$$
R \equiv R_{AB}g^{BA} = (-)^C\partial_C\Gamma^C_{AB}\cdot g^{BA} - (-)^B\partial_B(g^{BA}\Gamma_A) - g^{AB}\Gamma_B\Gamma_A - (-)^{(A+1)E}\Gamma^E_{AC}\Gamma^C_{EB}g^{BA}
$$

$$
= -g^{AB}\Gamma_B\Gamma_A - 2(-)^A\partial_A(g^{AB}\Gamma_B) - (-)^{A+B}\partial_A\partial_B g^{BA} + (-)^{C(A+1)}\Gamma^C_{AD}\Gamma^D_{CB}g^{BA} = \frac{(1-a)(2a-1)}{a^2}. \tag{A12}
$$

(f) Raising and lowering indices are defined as

$$\Phi^{A_1A_2\cdots A_n} = \Phi^{A_1A_2\cdots A_{n-1}}{}_{B_n} g^{B_nA_n} = (-)^{(A_{n-1}+B_{n-1})A_n}\Phi^{A_1\cdots A_{n-2}}{}_{B_{n-1}B_n} g^{B_nA_n} g^{B_{n-1}A_{n-1}}$$

$$\vdots$$

$$= (-)^{\sum_{k=1}^{n-1}\sum_{l=k+1}^{n}(B_k+A_k)A_l}\Phi_{B_1B_2\cdots B_n}\prod_{s=1}^{n} g^{B_sA_s}, \tag{A13}$$

we also have,

$$R_{ABCD} = (-)^A g_{AE} R^E{}_{BCD},$$
$$\Gamma_{ABC} = (-)^A g_{AD}\Gamma^D_{BC}, \tag{A14}$$
$$R^{AB}{}_{CD} = (-)^{(C+D)(B+E)} R^A{}_{ECD}\, g^{EB}.$$

(g) Symmetries:

$$g_{AB} = (-)^{A+B+AB} g_{BA}, \quad g^{AB} = (-)^{AB} g^{BA},$$
$$R^A{}_{BCD} = (-)^{CD} R^A{}_{BDC}, \quad R_{AB} = (-)^{AB} R_{BA},$$
$$R_{ABCD} = -(-)^{CD} R_{ABDC} = -(-)^{AB} R_{BACD}$$
$$= (-)^{(A+B)(C+D)} R_{CDAB}. \tag{A15}$$

(h) When $a = 1$, we find

$$g_{z\bar z} = \partial_z\,\partial_{\bar z}\log Y^{-2}, \quad g_{z\bar\theta} = -\partial_z\,\partial_{\bar\theta}\log Y^{-2},$$
$$g_{\theta\bar z} = \partial_\theta\,\partial_{\bar z}\log Y^{-2},$$
$$g_{\theta\bar\theta} = \partial_\theta\,\partial_{\bar\theta}\log Y^{-2}, \quad \text{others} = 0, \tag{A16}$$

and

$$\Gamma^z_{zz} = \frac{i}{Y}, \quad \Gamma^z_{z\theta} = \Gamma^z_{\theta z} = \frac{\bar\theta}{2Y}, \quad \Gamma^\theta_{z\theta} = \Gamma^\theta_{\theta z} = \frac{i}{2Y},$$
$$\Gamma^{\bar z}_{\bar z\bar z} = -\frac{i}{Y}, \quad \Gamma^{\bar z}_{\bar z\bar\theta} = \Gamma^{\bar z}_{\bar\theta\bar z} = -\frac{\theta}{2Y},$$
$$\Gamma^{\bar\theta}_{\bar z\bar\theta} = \Gamma^{\bar\theta}_{\bar\theta\bar z} = -\frac{i}{2Y},$$
$$\text{others} = 0. \tag{A17}$$

APPENDIX B: "SUPER" HYPERBOLIC DISTANCE $d(q_1,q_2)$

In this appendix we calculate the explicit form of the "super" hyperbolic distance $d(q_1,q_2)$. Plugging (59) into (55), we get the differential equation for $d(q_1,q_2)$,

$$g^{AB}(\partial_B d)(\partial_A d) = 1. \tag{B1}$$

Plugging (63) into the above equation we get

$$f^2 + 2fkr - 1 = g^{AB}(\partial_B R)(\partial_A R)\{(f')^2 + 2f'k'r\}$$
$$+ 2g^{AB}(\partial_B R)(\partial_A r)(f'k + k'kr)$$
$$+ g^{AB}(\partial_B r)(\partial_A r)k^2, \tag{B2}$$

where $f' = df/dR$, and $k' = dk/dR$. Due to the identities,

$$g^{AB}(\partial_B R)(\partial_A R) = R^2 + 4R + 4\,[(1-a)/a]Rr,$$
$$g^{AB}(\partial_B R)(\partial_A r) = (1/a)Rr, \tag{B3}$$
$$g^{AB}(\partial_B r)(\partial_A r) = -(1/a)r,$$

(B2) is reduced to two equations,

$$(f')^2(R^2+4R) = f^2 - 1, \tag{B4}$$

$$4[(1-a)/a](f')^2R + 2f'k'(R^2+4R)$$
$$+ (2/a)f'kR - (1/a)k^2 = 2fk. \tag{B5}$$

Equation $d(q,q) = 0$ induces a boundary condition for f,

$$f(0) = 1. \tag{B6}$$

Under this condition, (B4) can be immediately solved as,

$$f(R) = 1 + \tfrac{1}{2}R. \tag{B7}$$

Plugging (B7) into (B5) we have

$$k'(R^2+4R)$$
$$= (1/a)k^2 + ((a-1/a)R + 2)k + ((a-1)/a)R. \tag{B8}$$

We require that infinitesimal distance, $d^2(q,q+dq)$, is equal to the line element $ds^2 = dq^A g_{AB} q^B$. This gives a boundary condition for $k(R)$ such as

$$k(0) = -2a. \tag{B9}$$

We can rewrite (B8) and (B9) as

$$\frac{dy}{dx} = \frac{a^2-1}{a^2} + \frac{y}{2(x^2-x)} + \frac{y^2}{4(x^2-x)}, \tag{B10}$$

with

$$y(0) = 0, \tag{B11}$$

where

$$k(R) = -2a - ay(x) + 2(a-1)x,$$
$$R = -4x(>0). \tag{B12}$$

Equation (B10) is known as the Riccati's type differential equation. In order to solve that, we introduce a new variable $u(x)$ by,

$$y(x) = -\frac{4(x^2-x)}{u}\frac{du}{dx}. \tag{B13}$$

Then, the equation for $u(x)$ becomes a hypergeometric differential equation,

$$x(1-x)\frac{d^2u}{dx^2} + \{\gamma - (\alpha+\beta+1)x\}\frac{du}{dx} - \alpha\beta u = 0, \tag{B14}$$

where

$$\alpha = \tfrac{1}{2}(1 + 1/|a|), \quad \beta = 1 - \alpha, \quad \gamma = \tfrac{3}{2}. \tag{B15}$$

The solution of (B14) consistent with the boundary condition (B11) is

$$u(x) = (\text{const})\times F(\alpha, 1-\alpha; \tfrac{3}{2}; x), \tag{B16}$$

so that

$$y(x) = -\frac{4(x^2-x)}{F(\alpha,1-\alpha;\frac{3}{2};x)}\frac{d}{dx}F(\alpha,1-\alpha;\tfrac{3}{2};x). \tag{B17}$$

Due to formula (p. 1041 in Ref. 43),

$$F\left(\frac{1+\alpha}{2},\frac{1-\alpha}{2};\frac{3}{2};\sin^2 z\right) = \frac{\sin\alpha z}{(\sin z)}, \tag{B18}$$

we obtain the $k(R)$ in (65), and hence, we get (64).

APPENDIX C

We will give an outline of solving (66) with respect to the coordinates q and get the solutions of the Euler–Lagrange Eq. (51).

Let us call

$$q=(z,\bar{z},\theta,\bar{\theta}), \quad q_2=(w,\bar{w},\nu,\bar{\nu}),$$

$$\frac{\partial S}{\partial w} = -p, \quad \frac{\partial S}{\partial \bar{w}} = -\bar{p}, \quad \frac{\partial S}{\partial \nu} = i\bar{\xi}, \quad \frac{\partial S}{\partial \bar{\nu}} = i\xi. \tag{C1}$$

Here q_2, p, and ξ can be regarded as constants of integration. First we consider the case when p has a nonzero body part. We find that the equation,

$$\frac{\partial S/\partial w}{\partial S/\partial \bar{w}} = \frac{p}{\bar{p}} \equiv k^2, \quad \left(\bar{k} = \frac{1}{k}\right), \tag{C2}$$

gives

$$\left| z - \frac{kw+\bar{k}\bar{w}}{k+\bar{k}} - \frac{ik\nu\bar{\nu}}{k+\bar{k}} - i\bar{\nu}\theta + \frac{G+2}{2}\{(\nu+i\bar{\nu})\theta + i\nu\bar{\nu}\}\right|^2 = \left(\frac{2Y_{(2)}}{k+\bar{k}}\right)^2, \tag{C3}$$

where

$$G = G(d(q,q_2)) = \cosh d - 1 - \sinh d \coth\left(\frac{d}{2a}\right). \tag{C4}$$

Setting

$$C_1 \equiv \left|\frac{2Y_{(2)}}{k+\bar{k}}\right| \quad : \text{ real constant,}$$

$$C_2 \equiv \frac{kw+\bar{k}\bar{w}}{k+\bar{k}} + \frac{i(k-\bar{k})\nu\bar{\nu}}{2(k+\bar{k})} \quad : \text{ real constant,} \tag{C5}$$

we may rewrite (C3) as,

$$z = C_2 + (i/2)\nu\bar{\nu} + i\bar{\nu}\theta - [(G+2)/2] \times\{(\nu+i\bar{\nu})\theta + i\nu\bar{\nu}\} + C_1 e^{i\lambda}, \tag{C6}$$

where λ is a real function which will be determined later. The equation,

$$\frac{\partial S/\partial \nu}{\partial S/\partial w} = \frac{i\bar{\xi}}{\bar{p}} \equiv i(\bar{\eta}+\bar{\nu}), \tag{C7}$$

yields

$$\theta = \frac{(G+2)(e^{i\lambda}-i)\bar{\sigma} - iG(e^{i\lambda}+i)\sigma}{G^2(e^{i\lambda}+i)(e^{-i\lambda}-i) - (G+2)^2(e^{i\lambda}-i)(e^{-i\lambda}+i)}, \tag{C8}$$

where

$$\sigma \equiv (e^{-i\lambda}-i)(e^{i\lambda}-i)\eta - 2i\{(e^{-i\lambda}-i) + [(G+2)/2](e^{i\lambda}-e^{-i\lambda})\}\nu. \tag{C9}$$

Here and hereafter we set $k=\bar{k}=1$ which corresponds to choosing an arbitrary constant of the initial time t_0 properly and hence we can recover this degree of freedom by reintroducing t_0 to the solution later.

The above two equations (C6), (C8) yield, after some cumbersome calculations,

$$2Y\sinh d = C_1(e^{i\lambda}+e^{-i\lambda})\left[1 - \frac{G(e^{i\lambda}+i)(e^{-i\lambda}-i) + (G+2)(e^{i\lambda}-i)(e^{-i\lambda}+i)}{2C_1\mathscr{D}}\eta\bar{\eta} + \frac{G(G+2)}{2C_1\mathscr{D}}(e^{i\lambda}-e^{-i\lambda})(\eta-i\bar{\eta})(\nu+i\bar{\nu})\right], \tag{C10}$$

where

$$\mathscr{D} \equiv 2i(e^{i\lambda}-e^{-i\lambda})(G^2+2G+2) + 8(G+1). \tag{C11}$$

The third equation in (C1) yields

$$d = [p/(2am)]\{4aC_1 - \eta\bar{\eta} - i(a-1) \times(\eta-i\bar{\eta})(\nu+i\bar{\nu})\}t \equiv \omega t. \tag{C12}$$

Here ω is seen to be a real constant. Plugging (C12) into (C10), we get

$$e^{i\lambda} = \frac{\sinh d + i}{\cosh d}\left[1 - \frac{2i\sinh d}{C_1\cosh^2 d}(G+1-\cosh d)\eta\bar{\eta} + \frac{G(G+2)\tanh d}{C_1\mathscr{D}_0}\{(\eta-i\bar{\eta})(\nu+i\bar{\nu}) + i\tanh d(\eta+i\bar{\eta})(\nu+i\bar{\nu})\}\right], \tag{C13}$$

where

$$\mathscr{D}_0 \equiv 8(G+1) - (4/\cosh d)(G^2+2G+2)$$
$$= (4\sinh^2 d/\cosh d)(1-\coth[d/2a]). \quad \text{(C14)}$$

Hence from (C6), (C8), and (C13) we finally get the solution $(z^{(\mathrm{I})}(t),\theta^{(\mathrm{I})}(t))$ in (68) where we have reintroduced a constant t_0 or replaced d by X (71) and we have redefined the constants as,

$$\eta = 2(\xi_3 - \xi_1 + i\xi_2 - i\xi_4),$$
$$\nu = (1+i)(\xi_1 - i\xi_2) - (1-i)(\xi_3 - i\xi_4),$$
$$C_1 = c_1 - 2(\xi_1\xi_2 - \xi_1\xi_3 - \xi_1\xi_4 + \xi_2\xi_3 - \xi_2\xi_4 + \xi_3\xi_4),$$
$$C_2 = c_2 + 2\xi_1\xi_2. \quad \text{(C15)}$$

We can get the second solution $(z^{(\mathrm{II})}(t),\theta^{(\mathrm{II})}(t))$ (69) by redefining the constants as

$$\omega t_0 \to \omega t_0 - n, \quad \xi_1 \to (\xi_1/2)e^{[(a-1)/a]n},$$
$$\xi_2 \to \xi_2/2, \quad \xi_3 \to (\xi_3/2)e^{n/a}, \quad \xi_4 \to (\xi_4/2)e^{n},$$
$$c_1 + c_2 \to e^n, \quad c_1 - c_2 \to -c_2, \quad \text{(C16)}$$

and then taking the limit $n \to \infty$.

Secondly, we consider the case when the body part of p is zero. In this case we cannot take such a ratio as in (C2), however, we find that the third equation in (C1) yields

$$z(t) = \omega + \hat{z}(t), \quad \text{(C17)}$$

where $\hat{z}(t)$ does not have a body part, or $\hat{z}(t)$ takes a similar form as ω_s in (73) with the f_k's being functions of t and w. This simplifies Eqs. (66). For example, since

$$R = |z - w - \theta v|^2/YY_w \propto \nu\bar{\nu}\xi\bar{\xi}, \quad \text{(C18)}$$

we have

$$d^2 = -4ar + R,$$
$$G(d) = -2a - [(2a-1)(a-1)/6a]d^2, \quad \text{(C19)}$$

which means that d in this case does not have a body part either. The third equation in (C1) then yields

$$z = w + i\bar{\nu}(\theta - \nu) - (1-a)(\nu + i\bar{\nu})(\theta - \nu) + Y_w^2 pt + \frac{Y_w pt}{2}\Big[\frac{(\theta-\nu)(\bar{\theta}-\bar{\nu})}{3a} - iY_w^2 pt + (1-a)\{i(\nu+i\bar{\nu})(\theta-\nu) + (\nu+i\bar{\nu})(\bar{\theta}-\bar{\nu})\}\Big]. \quad \text{(C20)}$$

The fifth one yields

$$\theta = \nu + \eta t - \{ia\eta + (1-a)\bar{\eta}\} \times \frac{Y_w^2 p + i(1-a)(\nu + i\nu)\bar{\eta}}{2aY_w} t^2, \quad \text{(C21)}$$

where

$$\eta = -(i/2a)Y_w[\xi + \{a\nu - i(1-a)\bar{\nu}\}p]. \quad \text{(C22)}$$

Then from (C20) and (C21), we finally get the third solution $(z^{(\mathrm{III})}(t),\theta^{(\mathrm{III})}(t))$ in (70) where we have redefined the constants as,

$$\omega = ic_1 + c_2,$$
$$\omega_s = -(Y_w/2a)p + \{(1-a)\nu - ia\bar{\nu}\}\eta/(2aY_w),$$
$$\epsilon_1 = \eta, \quad \epsilon_2 = \nu. \quad \text{(C23)}$$

APPENDIX D: DERIVATION OF GRASSMANN EVEN EIGENFUNCTIONS (106)

We have found that the Grassmann even eigenfunctions for the operator $\square_0$ (98) takes the form

$$e_{(\lambda,k)}(Z) = C_{\lambda,k}(1 + (\lambda/2y)\theta\bar{\theta})e^{ikx}\sqrt{y}K_{\lambda-\frac{1}{2}}(|k|y). \quad \text{(D1)}$$

We show in this appendix how the normalization constants $C_{\lambda,k}$ are determined. The inner product of two eigenfunctions are calculated as,

$$\langle e_{(p,l)}|e_{(\lambda,k)}\rangle = \int_{-\infty}^{\infty}dx\int_0^{\infty}dy\int d\theta\int d\bar{\theta}\Big(\frac{1}{2ay + a\theta\bar{\theta}}\Big)\bar{e}_{(p,l)}(Z)e_{(\lambda,k)}(Z)$$
$$= \int_{-\infty}^{\infty}dx\int_0^{\infty}dy\,\bar{C}_{(p,l)}C_{(\lambda,k)}e^{i(k-l)x}\frac{1-\lambda-\bar{p}}{4ay}K_{\bar{p}-1/2}(|l|y)K_{\lambda-1/2}(|k|y)$$
$$= \frac{\pi(1-\lambda-\bar{p})}{2a}\bar{C}_{(p,l)}C_{(\lambda,k)}\delta(k-l)\int_0^{\infty}dy(1/y)K_{\bar{p}-1/2}(|k|y)K_{\lambda-1/2}(|k|y)$$
$$= \frac{\pi(1-\lambda-\bar{p})}{2a}\bar{C}_{(p,l)}C_{(\lambda,k)}\delta(k-l)I_{\lambda,\bar{p}}, \quad \text{(D2)}$$

where

$$I_{\lambda,\bar{p}} \equiv \int_0^{\infty}dz(1/z)\,K_{\bar{p}-1/2}(z)K_{\lambda-1/2}(z). \quad \text{(D3)}$$

Now we will evaluate $I_{\lambda,\bar{p}}$. Due to the asymptotic behavior of a modified Bessel function $K_\nu(z)$,

$$K_\nu(z) \sim \begin{cases} \sqrt{\pi/2z}\,e^{-z}(1+o(1)), & |z| \to \infty, \\ \frac{1}{2}\{(z/2)^\nu\Gamma(-\nu) + (z/2)^{-\nu}\Gamma(\nu)\}(1+o(1)), & z \to 0, \end{cases} \quad \text{(D4)}$$

we find that the condition.

$$1+|\mathrm{Re}(\rho-\tfrac{1}{2})|+|\mathrm{Re}(\lambda-\tfrac{1}{2})|<1, \tag{D5}$$

needs to be satisfied in order that the integral is well-defined. We then get (105) and the integral will give a δ function. In fact, due to a formula (p. 693 in Ref. 43),

$$\int_0^\infty dx\, x^{-\lambda}K_\mu(ax)K_\nu(bx)$$
$$=\frac{a^{\lambda-\nu-1}b^\nu}{2^{2+\lambda}\Gamma(1-\lambda)}\Gamma\Big(\frac{1-\lambda+\mu+\nu}{2}\Big)\Gamma\Big(\frac{1-\lambda-\mu+\nu}{2}\Big)\Gamma\Big(\frac{1-\lambda+\mu-\nu}{2}\Big)\Gamma\Big(\frac{1-\lambda-\mu-\nu}{2}\Big)$$
$$\times{}_2F_1\Big(\frac{1-\lambda+\mu+\nu}{2},\frac{1-\lambda-\mu+\nu}{2};\ 1-\lambda;1-b^2/a^2\Big),\quad [\mathrm{Re}(a+b)>1,\quad \mathrm{Re}\lambda<1-|\mathrm{Re}\,\mu|-|\mathrm{Re}\,\nu|\ ], \tag{D6}$$

we have $(pq>0)$

$$\int_0^\infty dx\, x^{-1}K_{ip}(x)K_{iq}(x)\equiv\lim_{\epsilon\to0}\int_0^\infty dx\, x^{\epsilon-1}K_{ip}(x)K_{iq}(x)$$
$$=\lim_{\epsilon\to0}\frac{\Gamma\Big(\frac{\epsilon+i(p+q)}{2}\Big)\Gamma\Big(\frac{\epsilon-i(p-q)}{2}\Big)\Gamma\Big(\frac{\epsilon+i(p-q)}{2}\Big)\Gamma\Big(\frac{\epsilon-i(p+q)}{2}\Big)}{2^{3-\epsilon}\Gamma(\epsilon)}$$
$$=\Big|\Gamma\Big(\frac{i(p+q)}{2}\Big)\Gamma\Big(1+\frac{i(p-q)}{2}\Big)\Big|^2\lim_{\epsilon\to0}\frac{\epsilon}{2\{\epsilon^2+(p-q)^2\}}$$
$$=\Big|\Gamma\Big(\frac{i(p+q)}{2}\Big)\Gamma\Big(1+\frac{(p-q)}{2}\Big)\Big|^2(\pi/2)\delta(p-q)$$
$$=\frac{\pi^2}{2p\sinh\pi p}\,\delta(p-q). \tag{D7}$$

From (D2), (D3), (D7) and due to symmetry, $K_\nu(z)=K_{-\nu}(z)$, we finally get the Grassmann even eigenfunctions (106).

APPENDIX E: DERIVATION OF GRASSMANN ODD EIGENFUNCTIONS (118)

The Grassmann odd eigenfunctions for the operator $\Box_0$ (98) were found to be

$$\psi_{(\lambda,k)}(Z)=C_{\lambda,k}(e^{ikx}/\sqrt{y})(\theta W_{\sigma_k/2,\lambda}(2|k|y)+i\bar\theta\lambda^{\sigma_k}W_{-\sigma_k/2,\lambda}(2|k|y)), \tag{E1}$$

where use has been made of a recursion formula of Whittaker functions (p. 1062 in Ref. 43),

$$z\partial_z W_{\kappa,\mu}(z)=(\kappa-z/2)W_{\kappa,\mu}(z)-[\mu^2-(\kappa-\tfrac{1}{2})^2]W_{\kappa-1,\mu}(z)=(z/2-\kappa)W_{\kappa,\mu}(z)-W_{\kappa+1,\mu}(z). \tag{E2}$$

The normalization constants $C_{\lambda,k}$ are determined as follows. The inner product of two eigenfunctions become

$$\langle\psi_{(\rho,l)}|\psi_{(\lambda,k)}\rangle=\int_{-\infty}^\infty dx\int_0^\infty dy\int d\theta\int d\bar\theta\Big(\frac{1}{2ay+a\theta\bar\theta}\Big)\bar\psi_{(\rho,l)}(Z)\psi_{(\lambda,k)}(Z)$$
$$=\int_{-\infty}^\infty dx\int_0^\infty dy\frac{\overline{C}_{(\rho,l)}C_{(\lambda,k)}}{2ay^2}e^{i(k-l)x}\{W_{\sigma_k/2,\bar\rho}(2|l|y)W_{\sigma_k/2,\lambda}(2|k|y)$$
$$-\rho^{-\sigma_l}\lambda^{\sigma_k}W_{-\sigma_k/2,\bar\rho}(2|l|y)W_{-\sigma_k/2,\lambda}(2|k|y)\}$$
$$=(\pi/a)\overline{C}_{(\rho,k)}C_{(\lambda,k)}|k|\delta(k-l)J^{\sigma_k}_{\lambda,\bar\rho}, \tag{E3}$$

where

$$J^{\sigma_k}_{\lambda,\bar\rho}=\int_0^\infty dz\,(1/z^2)\{W_{\sigma_k/2,\bar\rho}(2z)W_{\sigma_k/2,\lambda}(2z)-(\bar\rho\lambda)^{\sigma_k}W_{-\sigma_k/2,\bar\rho}(2z)W_{-\sigma_k/2,\lambda}(2z)\}\equiv\int_0^\infty dz\,\mathscr{J}^{\sigma_k}_{\lambda,\bar\rho}. \tag{E4}$$

Due to the asymptotic behavior of a Whittaker function,

$$W_{\kappa,\mu}(z)\sim\begin{cases} e^{-z/2}z^\kappa\Big(1+\frac{\mu^2-(\kappa-\frac{1}{2})^2}{z}+o(z^{-1})\Big), & |z|\to\infty,\\ \frac{\Gamma(-2\mu)}{\Gamma(\frac{1}{2}-\mu-\kappa)}z^{1/2+\mu}\Big(1-\frac{\kappa}{2\mu+1}z+o(z)\Big)+\frac{\Gamma(2\mu)}{\Gamma(\frac{1}{2}+\mu-\kappa)}z^{1/2-\mu}\Big(1-\frac{\kappa}{1-2\mu}z+o(z)\Big), & |z|\to0\end{cases} \tag{E5}$$

we find that of the integrand $\mathscr{J}^{\sigma_k}_{\lambda,\bar{\rho}}(z)$,

$$\mathscr{J}^{\sigma_k}_{\lambda,\bar{\rho}}(z) \sim \begin{cases} e^{-2z}\{(2z)^{\sigma_k} - (\bar{\rho}\lambda)^{\sigma_k}(2z)^{-\sigma_k}\}(1+o(1)),\ |z|\to\infty, \\ \Big[(2z)^{\bar{\rho}-\lambda-1}\dfrac{2\Gamma(-2\bar{\rho})\Gamma(2\lambda)}{\Gamma((1-\sigma_k)/2-\bar{\rho})\Gamma((1-\sigma_k)/2+\lambda)} \\ \quad + (2z)^{\lambda-\bar{\rho}-1}\dfrac{2\Gamma(2\bar{\rho})\Gamma(-2\lambda)}{\Gamma((1-\sigma_k)/2+\bar{\rho})\Gamma((1-\sigma_k)/2-\lambda)} \\ \quad - (2z)^{\bar{\rho}+\lambda}\dfrac{2\sigma_k(1+\bar{\rho}+\lambda)\Gamma(-2\bar{\rho})\Gamma(-2\lambda)}{(1+2\bar{\rho})(1+2\lambda)\Gamma((1-\sigma_k)/2-\bar{\rho})\Gamma((1-\sigma_k)/2-\lambda)} \\ \quad - 2(z)^{-\bar{\rho}-\lambda}\dfrac{2\sigma_k(1-\bar{\rho}-\lambda)\Gamma(2\bar{\rho})\Gamma(2\lambda)}{(1-2\bar{\rho})(1-2\lambda)\Gamma((1-\sigma_k)/2+\bar{\rho})\Gamma((1-\sigma_k)/2+\lambda)}\Big](1+o(1)),\ |z|\to 0. \end{cases} \tag{E6}$$

Thus λ and ρ should satisfy

$$1+|\mathrm{Re}(\bar{\rho}-\lambda)|<1, \quad \text{and} \quad |\mathrm{Re}(\bar{\rho}+\lambda)|<1, \tag{E7}$$

in order that the integral be well-defined. We then get (116) and (117) and $J^{\sigma_k}_{\lambda,\bar{\rho}}$ will give a δ function.

To evaluate $J^{\sigma_k}_{\lambda\bar{\rho}}$ we point out a relation in special functions. First we find that $z\{K_{\lambda+\frac{1}{2}}(z)+\sigma K_{\lambda-\frac{1}{2}}(z)\}$ satisfies the same second-order differential equation as $W_{\sigma/2,\lambda}(2z)$, where $\sigma=\pm 1$,

$$\{\partial_z^2+(-1+\sigma/z-(\lambda^2-1/4)/z^2)\}W_{\sigma/2,\lambda}(2z)=0.$$
$$\{\partial_z^2+(-1+\sigma/z-(\lambda^2-1/4)/z^2)\}[z\{K_{\lambda+\frac{1}{2}}(z)+\sigma K_{\lambda-\frac{1}{2}}(z)\}]=0. \tag{E8}$$

Comparing the asymptotic behavior for both $|z|\to\infty$ and $|z|\to 0$ of those two functions, we find that the following relation should exist;

$$W_{\sigma/2,\lambda}(2z)=\frac{\lambda^{(\sigma-1)/2}}{\sqrt{\pi}}z\{K_{\lambda+1/2}(z)+\sigma K_{\lambda-1/2}(z)\}, \quad (\sigma=\pm 1). \tag{E9}$$

Precisely speaking, the rhs should read

$$\lim_{\epsilon\to 0}\frac{(\lambda+\epsilon)^{(\sigma-1)/2}}{\sqrt{\pi}}z\{K_{\lambda+\epsilon+1/2}(z)+\sigma K_{\lambda+\epsilon-1/2}(z)\}, \tag{E10}$$

then, when $\sigma=-1$ and $\lambda=0$, we have

$$W_{-1/2,0}(2z)=\frac{2z}{\sqrt{\pi}}\frac{\partial K_\nu}{\partial \nu}\bigg|_{\nu=1/2}, \tag{E11}$$

which in fact holds. Due to another recursion formula of Whittaker functions (p. 1062 in Ref. 43),

$$W_{\kappa,\mu}(z)=z^{1/2}W_{\kappa-1/2,\mu\pm 1/2}(z)+(\tfrac{1}{2}-\kappa\pm\mu)W_{\kappa-1,\mu}(z), \tag{E12}$$

we find that the integrand $\mathscr{J}^{\sigma_k}_{\lambda,\bar{\rho}}$ becomes

$$\begin{aligned}\mathscr{J}^{\sigma_k}_{\lambda,\bar{\rho}} &= -4(-\lambda\bar{\rho})^{(\sigma_k-1)/2}[(2z)^{-1}W_{0,\bar{\rho}-1/2}(2z)W_{0,\lambda-1/2}(2z) \\ &\quad -(2z)^{-3/2}\{W_{1/2,\lambda}(2z)W_{0,\bar{\rho}-1/2}(2z)+W_{1/2,\bar{\rho}}(2z)W_{0,\lambda-1/2}(2z)\}] \\ &= (2/\pi)(-\lambda\bar{\rho})^{(\sigma_k-1)/2}\{K_{\bar{\rho}+1/2}(z)K_{\lambda-1/2}(z)+K_{\bar{\rho}-1/2}(z)K_{\lambda+1/2}(z)\}.\end{aligned} \tag{E13}$$

Plugging (E13) into (E4), we get

$$\begin{aligned}J^{\sigma_k}_{\lambda,\bar{\rho}} &= (2/\pi)(-\lambda\bar{\rho})^{(\sigma_k-1)/2}\int_0^\infty dz\{K_{\bar{\rho}+1/2}(z)K_{\lambda-1/2}(z)+K_{\bar{\rho}-1/2}(z)K_{\lambda+1/2}(z)\} \\ &\equiv (2/\pi)(-\lambda\bar{\rho})^{(\sigma_k-1)/2}\lim_{\epsilon\to 0}\int_0^\infty dz\, z^\epsilon\{K_{\bar{\rho}+1/2}(z)K_{\lambda-1/2}(z)+K_{\bar{\rho}-1/2}(z)K_{\lambda+1/2}(z)\} \\ &= (-\lambda\bar{\rho})^{(\sigma_k-1)/2}\lim_{\epsilon\to 0}\frac{2^{\epsilon-1}}{\pi\Gamma(1+\epsilon)}\Gamma\Big(\frac{1+\epsilon+\bar{\rho}+\lambda}{2}\Big)\Gamma\Big(\frac{1+\epsilon-\bar{\rho}-\lambda}{2}\Big) \\ &\quad\times\Big\{\Gamma\Big(\frac{\epsilon-\bar{\rho}-\lambda}{2}\Big)\Gamma\Big(\frac{2+\epsilon+\bar{\rho}-\lambda}{2}\Big)+\Gamma\Big(\frac{2+\epsilon-\bar{\rho}+\lambda}{2}\Big)\Gamma\Big(\frac{\epsilon+\bar{\rho}-\lambda}{2}\Big)\Big\}\end{aligned}$$

$$= (- \lambda \bar{\rho})^{(\sigma_k - 1)/2} \lim_{\epsilon \to 0} \frac{2^{\epsilon - 1}}{\pi \Gamma(1+\epsilon)} \Gamma\Big(\frac{1+\epsilon+\bar{\rho}+\lambda}{2}\Big)\Gamma\Big(\frac{1+\epsilon-\bar{\rho}-\lambda}{2}\Big)\Gamma\Big(\frac{2+\epsilon-\bar{\rho}+\lambda}{2}\Big)$$
$$\times \Gamma\Big(\frac{2+\epsilon+\bar{\rho}-\lambda}{2}\Big)\Big(\frac{2}{\epsilon-\bar{\rho}+\lambda}+\frac{2}{\epsilon+\bar{\rho}-\lambda}\Big)$$
$$= \frac{(-\lambda\bar{\rho})^{(\sigma_k-1)/2}}{2\pi} \Gamma\Big(\frac{1+\bar{\rho}+\lambda}{2}\Big)\Gamma\Big(\frac{1-\bar{\rho}-\lambda}{2}\Big)\Gamma\Big(\frac{2-\bar{\rho}+\lambda}{2}\Big)\Gamma\Big(\frac{2+\bar{\rho}-\lambda}{2}\Big)\sum_{\epsilon\to 0}^{\lim} \frac{4\epsilon}{\epsilon^2-(\lambda-\bar{\rho})^2}, \tag{E14}$$

where use has been made of the formula (D2). According to (116) we will rewrite the variables as

$$\lambda = c + ip, \quad \rho = c + iq. \tag{E15}$$

Then we finally get

$$J^{\sigma_k}_{c+ip,c+iq} = \frac{2\pi\sigma_k (c+ip)^{\sigma_k - 1}}{\cos[\pi(c+ip)]}\delta(p+q). \tag{E16}$$

Hence, from (E3) and (E16) we get the Grassmann odd eigenfunctions (118).

APPENDIX F: DERIVATION OF THE COMPLETENESS RELATION (123)

First we evaluate the contribution of the Grassmann even eigenfunctions,

$$I^B(q_1,q_2) \equiv \int_{-\infty}^{\infty} dp\, dk\, e_{p,k}(q_2)\bar{e}_{-p,k}(q_1). \tag{F1}$$

Plugging (106) into (F1), we get

$$I^B(q_1,q_2) = \int_{-\infty}^{\infty} dp\, dk \frac{2ia}{\pi^3}\sqrt{y_1y_2}e^{ik(x_2-x_1)}\Big(1+\frac{1+2ip}{4y_1}\theta_1\bar{\theta}_1\Big)\Big(1+\frac{1+2ip}{4y_2}\theta_2\bar{\theta}_2\Big)\sinh \pi p\, K_{ip}(|k|y_1)K_{ip}(|k|y_2)$$
$$= -\int_{-\infty}^{\infty} dp\, dk \frac{2a\sqrt{y_1y_2}}{\pi^3}\Big(\frac{\theta_1\bar{\theta}_1}{2y_1}+\frac{\theta_2\bar{\theta}_2}{2y_2}+\frac{\theta_1\bar{\theta}_1\theta_2\bar{\theta}_2}{4y_1y_2}\Big)e^{ik(x_2-x_1)}K_{ip}(|k|y_1)K_{ip}(|k|y_2). \tag{F2}$$

Differentiating a formula (p. 510 in Ref. 43),

$$\int_0^{\infty} dx \cosh[(\pi-\phi)x]K_{ix}(a)K_{ix}(b) = (\pi/2)K_0(\sqrt{a^2+b^2-2ab\cos\phi}) \tag{F3}$$

with respect to ϕ, we get[44]

$$\int_0^{\infty} dx\, x \sinh[(\pi-\phi)x]K_{ix}(a)K_{ix}(b) = \frac{\pi ab \sin\phi}{2\sqrt{a^2+b^2-2ab\cos\phi}}K_1(\sqrt{a^2+b^2-2ab\cos\phi}). \tag{F4}$$

Considering that $K_1 \sim x^{-1} + o(1)$, $x \to 0$, we take the limit of $\phi \to 0$,[44]

$$\int_0^{\infty} dx\, x \sinh \pi x\, K_{ix}(a)K_{ix}(b) = \lim_{\phi\to 0}\frac{\pi ab\sin\phi}{2(a^2+b^2-2ab\cos\phi)} = (\pi^2/2)\sqrt{ab}\,\delta(a-b). \tag{F5}$$

Then $I^B(q_1,q_2)$ is obtained,

$$I^B(q_1,q_2) = -2ay_1\Big(\theta_1\bar{\theta}_1+\theta_2\bar{\theta}_2+\frac{\theta_1\bar{\theta}_1\theta_2\bar{\theta}_2}{2y_1}\Big)\delta(x_1-x_2)\delta(y_1-y_2). \tag{F6}$$

Next we consider the contribution of the Grassmann odd part,

$$I^F_c(q_1,q_2) \equiv \int_{-\infty}^{\infty} dp\, dk\, \psi^c_{p,k}(q_2)\bar{\psi}^c_{-p,k}(q_1)$$
$$= \int_{-\infty}^{\infty} dp\, dk \frac{a\cos[\pi(c+ip)]}{2\pi^2 k}\frac{e^{ik(x_2-x_1)}}{\sqrt{y_1y_2}}\big[\theta_2\bar{\theta}_1(c+ip)^{1-\sigma_k}W_{\sigma_k/2,c+ip}(2|k|y_1)W_{\sigma_k/2,c+ip}(2|k|y_2)$$
$$+ \bar{\theta}_2\theta_1(c+ip)^{1+\sigma_k}W_{-\sigma_k/2,c+ip}(2|k|y_1)W_{-\sigma_k/2,c+ip}(2|k|y_2)$$
$$- i\bar{\theta}_1\bar{\theta}_2(c+ip)W_{\sigma_k/2,c+ip}(2|k|y_1)W_{-\sigma_k/2,c+ip}(2|k|y_2)$$
$$+ i\theta_1\theta_2(c+ip)W_{-\sigma_k/2,c+ip}(2|k|y_1)W_{\sigma_k/2,c+ip}(2|k|y_2)\big]$$
$$= \int_{-\infty}^{\infty} dk \frac{a\sqrt{y_1y_2}}{2\pi^3}ke^{ik(x_2-x_1)}[\theta_2\bar{\theta}_1\mathscr{I}^{++}_c(k)+\bar{\theta}_2\theta_1\mathscr{I}^{--}_c(k)-i\bar{\theta}_1\bar{\theta}_2\mathscr{I}^{+-}_c(k)+i\theta_1\theta_2\mathscr{I}^{-+}_c(k)], \tag{F7}$$

where

$$\mathscr{I}_c^{\epsilon_1\epsilon_2}(k) \equiv \int_{-\infty}^{\infty} dp \cos[\pi(c+ip)]\{K_{c+ip+1/2}(|k|y_1) + \epsilon_1\sigma_k K_{c+ip-1/2}(|k|y_1)\}$$
$$\times\{K_{c+ip+1/2}(|k|y_2) + \epsilon_2\sigma_k K_{c+ip-1/2}(|k|y_2)\}, \quad (\epsilon_1,\epsilon_2 = \pm). \tag{F8}$$

Here use has been made of (E9). Due to the symmetry $K_\nu(z) = K_{-\nu}(z)$, $\mathscr{I}_c^{\epsilon_1\epsilon_2}(k)$ satisfies

$$\mathscr{I}_{-c}^{\epsilon_1\epsilon_2}(k) = \epsilon_1\epsilon_2\mathscr{I}_c^{\epsilon_1\epsilon_2}(k). \tag{F9}$$

We will evaluate $\mathscr{I}_c^{\epsilon_1\epsilon_2}(k)$ at $c=0$ first. Equation (F9) yields

$$\mathscr{I}_0^{+-}(k) = \mathscr{I}_0^{-+}(k) = 0. \tag{F10}$$

Consider a formula (p.771 in Ref. 43),[45]

$$\int_{-\infty}^{\infty} dx\, e^{ipx} K_{\nu+ix}(\alpha)K_{\nu-ix}(\beta)$$
$$= \pi\left(\frac{\alpha e^{\rho}+\beta}{\alpha+\beta e^{\rho}}\right)^{\nu} K_{2\nu}(\sqrt{\alpha^2+\beta^2+2\alpha\beta\cosh\rho}),$$
$$\times[|\arg\alpha| + |\arg\beta| + |\mathrm{Im}\,\rho| < \pi]. \tag{F11}$$

Using recursion formulas of modified Bessel functions (pp. 267, 968 in Ref. 43),

$$\begin{aligned} z\partial_z K_\nu(z) + \nu K_\nu(z) &= -zK_{\nu-1}(z),\\ z\partial_z K_\nu(z) - \nu K_\nu(z) &= -zK_{\nu+1}(z), \end{aligned} \tag{F12}$$

we have

$$\begin{aligned} &\int_{-\infty}^{\infty} dx \cosh[(\pi-\phi)x]K_{\frac{1}{2}+ix}(a)K_{1/2-ix}(b)\\ &\quad= \frac{\pi(a+b)\sin(\phi/2)}{\sqrt{a^2+b^2-2ab\cos\phi}}K_1(\sqrt{a^2+b^2-2ab\cos\phi}),\\ &\int_{-\infty}^{\infty} dx \cosh[(\pi-\phi)x]K_{\frac{1}{2}+ix}(a)K_{\frac{1}{2}+ix}(b)\\ &\quad= \pi\sin(\phi/2)\, K_0(\sqrt{a^2+b^2-2ab\cos\phi}),\\ &\int_{-\infty}^{\infty} dx \cosh[(\pi-\phi)x]K_{\frac{1}{2}-ix}(a)K_{\frac{1}{2}-ix}(b)\\ &\quad= \pi\sin(\phi/2)K_0(\sqrt{a^2+b^2-2ab\cos\phi}). \end{aligned} \tag{F13}$$

Due to the asymptotic behaviors of K_0 and K_1,

$$\begin{aligned} K_0(z) &\sim -\log z + O(1),\\ K_1(z) &\sim 1/z + o(1), \end{aligned} \tag{F14}$$

we get

$$\begin{aligned} &\int_{-\infty}^{\infty} dx \cosh(\pi x)\, K_{1/2+ix}(a)K_{1/2-ix}(b) = \pi^2\delta(a-b),\\ &\int_{-\infty}^{\infty} dx \cosh(\pi x)\, K_{1/2+ix}(a)K_{1/2+ix}(b)\\ &\quad= \int_{-\infty}^{\infty} dx \cosh(\pi x)\, K_{1/2-ix}(a)K_{1/2-ix}(b) = 0. \end{aligned} \tag{F15}$$

Hence we have

$$\mathscr{I}_c^{\epsilon_1\epsilon_2}(k) = \frac{\pi^2(\epsilon_1+\epsilon_2)}{k}\delta(y_1-y_2). \tag{F16}$$

Next we will evaluate $\mathscr{I}_c^{\epsilon_1\epsilon_2}(k)$ at $c=\frac{1}{2}$. Due to (F11), $\mathscr{I}_{1/2}^{\epsilon_1\epsilon_2}(k)$ is rewritten as

$$\mathscr{I}_{1/2}^{\epsilon_1\epsilon_2}(k) = \left\{\frac{1}{k}\left(\frac{\epsilon_1}{y_2}+\frac{\epsilon_2}{y_1}\right) - \frac{1}{k^2}\left(\frac{1}{y_1}\frac{\partial}{\partial y_2}+\frac{1}{y_2}\frac{\partial}{\partial y_1}\right)\right\}$$
$$\times\int_{-\infty}^{\infty} dp\, p\sinh(\pi p)\, K_{ip}(|k|y_1)K_{ip}(|k|y_2). \tag{F17}$$

Plugging (F5), we finally get

$$\mathscr{I}_{1/2}^{\epsilon_1\epsilon_2}(k) = [\pi^2(\epsilon_1+\epsilon_2)/k]\delta(y_1-y_2). \tag{F18}$$

Equations (F16) and (F18) implies

$$\mathscr{I}_c^{\epsilon_1\epsilon_2}(k) = [\pi^2(\epsilon_1+\epsilon_2)/k]\delta(y_1-y_2), \quad 0 \leqslant c \leqslant \tfrac{1}{2}. \tag{F19}$$

In fact, a relation (p. 970 in Ref. 43),

$$K_\nu(z) = \frac{\pi}{2}\frac{I_{-\nu}-I_\nu}{\sin(\nu\pi)}, \quad [\nu\text{: not an integer}], \tag{F20}$$

implies that $K_{ip}(z)$ is analytic in p where $|\mathrm{Im}\,p| < 1$, and hence we may change the path of integration in $\mathscr{I}_0^{\epsilon_1\epsilon_2}(k)$ from $(-\infty,\infty)$ to $(-\infty-ic,\infty-ic)$, which gives $\mathscr{I}_c^{\epsilon_1\epsilon_2}(k)$. Due to (F8) we conclude,

$$\mathscr{I}_c^{\epsilon_1\epsilon_2}(k) = [\pi^2(\epsilon_1+\epsilon_2)/k]\delta(y_1-y_2), \quad |c| \leqslant \tfrac{1}{2}. \tag{F21}$$

We then get

$$I_c^F(q_1,q_2) = 2a\sqrt{y_1y_2}(\theta_1\bar\theta_2 - \bar\theta_1\theta_2)\delta(x_1-x_2)\delta(y_1-y_2). \tag{F22}$$

$I^B(q_1,q_2)$ and $I_c^F(q_1,q_2)$ gives (123).

APPENDIX G: DERIVATION OF THE HEAT KERNEL (128)

First, we evaluate the contribution of the Grassmann even part,

$$K^B(q_1,q_2|\tau) \equiv \int_{-\infty}^{\infty} dp\, dk\, e^{-\tau\gamma^B(p)} e_{p,k}(q_2)\bar e_{p,k}(q_1). \tag{G1}$$

Plugging (106) and (109) we get [cf. (F2)]

$$\begin{aligned} &K^B(q_1,q_2|\tau)\\ &\quad= \int_{-\infty}^{\infty} dp \frac{2ia}{\pi^3}\sqrt{y_1y_2}\exp[-\tau(p-i/2a)\\ &\qquad-\tau((1-a)/2a)^2]\left(1+\frac{1+2ip}{4y_1}\theta_1\bar\theta_1\right)\\ &\qquad\times\left(1+\frac{1+2ip}{4y_2}\theta_2\bar\theta_2\right)\mathscr{F}(p), \end{aligned} \tag{G2}$$

where

$$\mathscr{F}(p) \equiv \int_{-\infty}^{\infty} dk \sinh(\pi p)\, e^{ik(x_2-x_1)} K_{ip}(|k|y_1) K_{ip}(|k|y_2)$$
$$= 2\sinh(\pi p) \int_0^{\infty} dk \cos[k(x_2-x_1)] \times K_{ip}(ky_1) K_{ip}(ky_2). \tag{G3}$$

Due to a formula (p. 732 in Ref. 43),

$$\int_0^{\infty} dx \cos(cx)\, K_\nu(ax) K_\nu(bx) = \frac{\pi^2}{4\sqrt{ab}} \sec(\nu\pi)\, P_{\nu-1/2}\left(\frac{a^2+b^2+c^2}{2ab}\right),$$
$$[\mathrm{Re}(a+b)>0, \quad c>0, \quad |\mathrm{Re}\,\nu| < \tfrac{1}{2}], \tag{G4}$$

we find

$$\mathscr{F}(p) = \frac{\pi^2}{2\sqrt{y_1 y_2}} \tanh(\pi p)\, P_{ip-1/2}(\cosh l_0), \tag{G5}$$

where [see (60)]

$$\cosh l_0 = 1 + \frac{(x_1-x_2)^2 + (y_1-y_2)^2}{2y_1 y_2} \equiv 1 + \tfrac{1}{2} R_0(q_1,q_2). \tag{G6}$$

A functional relation (p. 1020 in Ref. 43),

$$Q_\nu(z) - Q_{-\nu-1}(z) = \pi \cot(\nu\pi)\, P_\nu(z) \quad [\sin(\nu\pi) \neq 0], \tag{G7}$$

and an integral representation of a Legendre function (p. 1002 in Ref. 43),

$$Q_\nu(\cosh\alpha) = \frac{1}{\sqrt{2}} \int_\alpha^{\infty} db \frac{e^{-(\nu+1/2)b}}{(\cosh b - \cosh\alpha)^{1/2}}$$
$$[\alpha>0, \quad \mathrm{Re}\,\nu > -1], \tag{G8}$$

leads to,

$$\pi \tanh(\pi p)\, P_{ip-1/2}(\cosh\alpha) = \sqrt{2} \int_\alpha^{\infty} db \frac{\sin pb}{(\cosh b - \cosh\alpha)^{1/2}}, \tag{G9}$$

and hence we get

$$\mathscr{F}(p) = \frac{\pi}{\sqrt{2y_1y_2}} \int_{l_0}^{\infty} db \frac{\sin pb}{(\cosh b - \cosh l_0)^{1/2}}. \tag{G10}$$

Then $K^B(q_1,q_2|\tau)$ becomes

$$K^B(q_1,q_2|\tau) = \frac{\sqrt{2}ia}{\pi^2} \int_{l_0}^{\infty} db \frac{e^{-((1-a)/2a)^2\tau}}{(\cosh b - \cosh l_0)^{1/2}} \times \int_{-\infty}^{\infty} dp\, e^{-(p-i/2a)^2\tau}$$
$$\times \sin(pb) \left(1 + \frac{1+2ip}{4y_1} \theta_1\bar\theta_1\right) \times \left(1 + \frac{1+2ip}{4y_2} \theta_2\bar\theta_2\right). \tag{G11}$$

The integration with respect to p can be easily done and due to the identity,

$$\frac{e^{-b^2/4\tau}[(1-a^2)/a^2 - 2/\tau + b^2/\tau^2]\sinh(b/2a) - (2b/a\tau)\cosh(b/2a)}{(\cosh b - \cosh l_0)^{1/2}}$$
$$= \frac{8\cosh l_0\, (d/db)(e^{-b^2/4\tau}\sinh(b/2a)/\sinh b) + 4(\cosh^2 l_0 - 1)(d/db)\{(1/\sinh b)(d/db)(e^{-b^2/4\tau}\sinh(b/2a)/\sinh b)\}}{(\cosh b - \cosh l_0)^{1/2}}$$
$$+ \frac{d}{db}\left[(\cosh b - \cosh l_0)^{1/2} \left\{\frac{4(\cosh b + \cosh l_0)}{\sinh b} \frac{d}{db}\left(e^{-b^2/4\tau} \frac{\sinh(b/2a)}{\sinh b}\right) + 6e^{-b^2/4\tau} \frac{\sinh(b/2a)}{\sinh b}\right\}\right], \tag{G12}$$

we finally obtain

$$K^B(q_1,q_2|\tau) = -\frac{2ae^{-((1-a)/2a)^2\tau}}{\pi\sqrt{2\pi\tau}} \int_{l_0}^{\infty} db \frac{1}{(\cosh b - \cosh l_0)^{1/2}} \left[e^{-b^2/4\tau} \sinh\frac{b}{2a}\right.$$
$$+ e^{-b^2/4\tau}\left(\frac{a-1}{a}\sinh\frac{b}{2a} + \frac{b}{\tau}\cosh\frac{b}{2a}\right)\left(\frac{\theta_1\bar\theta_1}{4y_1} + \frac{\theta_2\bar\theta_2}{4y_2} + \frac{\theta_1\bar\theta_1\theta_2\bar\theta_2}{8y_1y_2}\right)$$
$$+ 2\cosh l_0 \frac{d}{db}\left(e^{-b^2/4\tau}\frac{\sinh(b/2a)}{\sinh b}\right) \frac{\theta_1\bar\theta_1\theta_2\bar\theta_2}{4y_1y_2}$$
$$\left. + (\cosh^2 l_0 - 1)\frac{d}{db}\left\{\frac{1}{\sinh b}\frac{d}{db}\left(e^{-b^2/4\tau}\frac{\sinh(b/2a)}{\sinh b}\right)\right\} \frac{\theta_1\bar\theta_1\theta_2\bar\theta_2}{4y_1y_2}\right]. \tag{G13}$$

Next we consider the Grassmann odd contribution,

$$K_c^F(q_1,q_2|\tau) \equiv \int_{-\infty}^{\infty} dp\, dk\, e^{-\tau\gamma_c^F(p)} \psi_{p,k}^c(q_2) \bar\psi_{-p,k}^c(q_1)$$
$$= \int_{-\infty}^{\infty} dp \frac{a\sqrt{y_1y_2}\cos\pi(c+ip)}{2\pi^3} \exp[-\tau(p-ic-i/2a)^2 - \tau((1-a)/2a)^2]$$
$$\times[(\bar\theta_2\theta_1 - \bar\theta_1\theta_2)G^+(p) + i(\theta_1\theta_2 - \bar\theta_1\bar\theta_2)G^-(p) - (\bar\theta_1\theta_2 + \bar\theta_2\theta_1)H^+(p) + i(\theta_1\theta_2 + \bar\theta_1\bar\theta_2)H^-], \tag{G14}$$

where

$$G^{\pm}(p) \equiv \int_{-\infty}^{\infty} dk\, k e^{ik(x_2-x_1)} [K_{c+ip+1/2}(|k|y_1) K_{c+ip+1/2}(|k|y_2) \pm K_{c+ip-1/2}(|k|y_1) K_{c+ip-1/2}(|k|y_2)], \tag{G15}$$

$$H^{\pm}(p) \equiv \int_{-\infty}^{\infty} dk\, k e^{ik(x_2-x_1)} \sigma_k [K_{c+ip+1/2}(|k|y_1) k_{c+ip-1/2}(|k|y_2) \pm K_{c+ip-1/2}(|k|y_1) K_{c+ip+1/2}(|k|y_2)]. \tag{G16}$$

Here use has been made of (E9) [cf. (F7)]. Due to a formula (p. 743 in Ref. 43),

$$\int_0^{\infty} dx\, x \sin(cx)\, K_\nu(ax) K_\nu(bx) = \frac{\pi c}{4(ab)^{3/2}} \left\{ \frac{(a-b)^2 + c^2}{2ab} \right\}^{-1/2} \Gamma(\tfrac{3}{2}+\nu)\Gamma(\tfrac{3}{2}-\nu) P_{\nu-1/2}^{-1}\left(\frac{a^2+b^2+c^2}{2ab} \right),$$
$$[\mathrm{Re}(a+b) > 0, \quad c > 0, \quad |\mathrm{Re}\,\nu| < \tfrac{3}{2}], \tag{G17}$$

and recursion formulas (p. 126 in Ref. 46),

$$P_{\nu-1}^{\mu}(z) - zP_\nu^\mu + (\nu - \mu + 1)\sqrt{z^2-1} P_\nu^{\mu-1}(z) = 0,$$
$$zP_\nu^\mu - P_{\nu+1}^\mu(z) + (\nu+\mu)\sqrt{z^2-1} P_\nu^{\mu-1}(z) = 0, \tag{G18}$$

we find

$$G^{\pm}(p) = \frac{i\pi(x_2-x_1)\Gamma(1+c+ip)\Gamma(1-c-ip)}{2(y_1y_2)^{3/2}(\cosh^2 l_0 - 1)^{1/2}} \{(1+c+ip) P_{c+ip}^{-1}(\cosh l_0) \pm (1-c-ip) P_{c+ip-1}^{-1}(\cosh l_0)\}$$
$$= \frac{i\pi^2(x_2-x_1)(c+ip)\{P_{c+ip}(\cosh l_0) \pm P_{-c-ip}(\cosh l_0)\}}{2(y_1y_2)^{3/2}(\cosh l_0 \pm 1)\sin[\pi(c+ip)]}, \tag{G19}$$

where use has been made of a symmetry, $P_\nu(z) = P_{-\nu-1}(z)$.

Due to (G4) and recursion formulas (F12) and (p. 1019 in Ref. 43),

$$(z^2-1)\partial_z P_\nu(z) = (\nu+1)\{P_{\nu+1}(z) - zP_\nu(z)\}, \tag{G20}$$

we get similarly[47]

$$H^{\pm}(p) = \frac{\pi^2(c+ip)(y_2 \pm y_1)\{P_{c+ip}(\cosh l_0) \pm P_{-c-ip}(\cosh l_0)\}}{2(y_1y_2)^{3/2}(\cosh l_0 \pm 1)\sin[\pi(c+ip)]}. \tag{G21}$$

So far $K_c^F(q_1,q_2|\tau)$ has been calculated as

$$K_c^F(q_1,q_2|\tau) = \frac{a}{4\pi^2} \exp[-((1-a)/2)^2\tau] \left\{ \frac{i(z_1-\bar{z}_2)\bar{\theta}_1\theta_2 - i(\bar{z}_1 - z_2)\bar{\theta}_2\theta_1}{y_1y_2} \mathscr{K}_c^{+}(l_0) + \frac{(\bar{z}_1-\bar{z}_2)\theta_1\theta_2 - (z_1-z_2)\bar{\theta}_1\bar{\theta}_2}{y_1y_2} \mathscr{K}_c^{-}(l_0) \right\}, \tag{G22}$$

where

$$\mathscr{K}_c^{\pm}(l_0) \equiv \int_{-\infty}^{\infty} dp \exp[-\tau(p - ic - i/2a)^2] \frac{\pi(c+ip)\cot[\pi(c+ip)]}{\cos l_0 \pm 1} \{P_{c+ip}(\cosh l_0) \pm P_{-c-ip}(\cosh l_0)\}. \tag{G23}$$

Equation (G7) and a recursion formula (p. 1019 in Ref. 43),

$$(2\nu+1) z Q_\nu(z) = (\nu+1) Q_{\nu+1}(z) + \nu Q_{\nu-1}(z), \tag{G24}$$

yield

$$\pi\nu \cot(\pi\nu) P_\nu(z) = \nu Q_\nu(z) + (\nu-1) Q_{1-\nu}(z) - (2\nu-1) z Q_{-\nu}(z),$$
$$\pi\nu \cot(\pi\nu) P_{-\nu}(z) = (2\nu+1) z Q_\nu(z) - (\nu+1) Q_{\nu+1}(z) - \nu Q_{-\nu}(z). \tag{G25}$$

Then using the integral representation (G8) we get

$$\mathcal{K}_c^{\pm}(l_0)=\sqrt{\frac{\pi}{2\tau}}\int_{l_0}^{\infty}dt\,\frac{1}{\cosh l_0\pm1}\frac{e^{-b^2/4\tau}}{(\cosh b-\cosh l_0)^{1/2}}\Bigg[e^{b/2a-b}\left\{\frac{b}{2\tau}-\frac{1-a}{2a}\pm\left(\frac{2a-1}{a}+\frac{b}{\tau}\right)\cosh l_0\right\}$$
$$+e^{-b/2a}\left\{\left(\frac{1}{a}+\frac{b}{\tau}\right)\cosh l_0\pm\left(\frac{1-a}{2a}+\frac{b}{2\tau}\right)\right\}$$
$$-e^{-b/2a-b}\left(\frac{a+1}{2a}+\frac{b}{2\tau}\right)\mp e^{b/2a-2b}\left(\frac{3a-1}{2a}+\frac{b}{2\tau}\right)\Bigg]$$
$$=\sqrt{\frac{\pi}{2\tau}}\int_{l_0}^{\infty}db\,\frac{2}{(\cosh b-\cosh l_0)^{1/2}}\Bigg[(\cosh l_0\mp1)\frac{d}{db}\left(e^{-b^2/4\tau}\frac{\sinh(b/2a)}{\sinh b}\right)$$
$$+e^{-b^2/4\tau}\left(\frac{b}{2\tau}\cosh\frac{b}{2a}+\frac{a-1}{2a}\sinh\frac{b}{2a}\right)\Bigg],\tag{G26}$$

where use has been made of identities,

$$\frac{e^{-b^2/4\tau}[e^{b/2a-b}((2a-1)/a+b/\tau)\cosh l_0-e^{b/2a-2b}((3a-1)/2a+b/2\tau)+e^{-b/2a}((1-a)/2a+b/2\tau)]}{(\cosh b-\cosh l_0)^{1/2}}$$
$$=\frac{e^{-b^2/4\tau}((b/\tau)\cosh(b/2a)+[(a-1)/a]\sinh(b/2a))}{(\cosh b-\cosh l_0)^{1/2}}+2\frac{d}{db}[(\cosh b-\cosh l_0)^{1/2}e^{-b^2/4\tau+b/2a-b}],\tag{G27}$$

$$\frac{e^{-b^2/4\tau}[e^{b/2a-b}(b/2\tau-(1-a)/2a)-e^{-b/2a-2b}((a+1)/2a+b/2\tau)+e^{-b/2a}(1/a+b/\tau)\cosh l_0]}{(\cosh b-\cosh l_0)^{1/2}}$$
$$=\frac{2(\cosh^2 l_0-1)\dfrac{d}{db}\left(e^{-b^2/4\tau}\dfrac{\sinh(b/2a)}{\sinh b}\right)+\cosh l_0\,e^{-b^2/4\tau}\left(\dfrac{b}{\tau}\cosh\dfrac{b}{2a}+\dfrac{a-1}{a}\sinh\dfrac{b}{2a}\right)}{(\cosh b-\cosh l_0)^{1/2}}$$
$$+2\frac{d}{db}\left[(\cosh b-\cosh l_0)^{1/2}(\cosh b+\cosh l_0-1)\,e^{-b^2/4\tau}\frac{\sinh(b/2a)}{\sinh b}\right].\tag{G28}$$

Plugging (G26) into (G22) we find

$$K_c^F(q_1,q_2|\tau)=-\frac{2ae^{-((1-a)/(2a))^2\tau}}{\pi\sqrt{2\pi\tau}}\int_{l_0}^{\infty}db\,\frac{1}{(\cosh b-\cosh l_0)^{1/2}}$$
$$\times\Bigg[\left\{\frac{\Delta R}{2}+(\cosh l-1)r-\cosh l\,\frac{\theta_1\bar\theta_1\theta_2\bar\theta_2}{2y_1y_2}\right\}\frac{d}{db}\left(e^{-b^2/4\tau}\frac{\sinh(b/2a)}{\sinh b}\right)$$
$$+\left\{r-\left(\frac{\theta_1\bar\theta_1}{2y_1}+\frac{\theta_2\bar\theta_2}{2y_2}+\frac{\theta_1\bar\theta_1\theta_2\bar\theta_2}{4y_1y_2}\right)\right\}e^{-b^2/4\tau}\left(\frac{b}{2\tau}\cosh\frac{b}{2a}+\frac{a-1}{2a}\sinh\frac{b}{2a}\right)\Bigg],\tag{G29}$$

where $r=r(q_1,q_2)$ and $l=l(q_1,q_2)$ are given in (62) and (65), respectively, and

$$\Delta R\equiv R(q_1,q_2)-R_0(q_1,q_2)=\frac{-(\bar z_1-\bar z_2)\theta_1\theta_2+(z_1-z_2)\bar\theta_1\bar\theta_2-\theta_1\bar\theta_1\theta_2\bar\theta_2}{y_1y_2}-2(\cosh l-1)\left(\frac{\theta_1\bar\theta_1}{2y_1}+\frac{\theta_2\bar\theta_2}{2y_2}+\frac{\theta_1\bar\theta_1\theta_2\bar\theta_2}{4y_1y_2}\right).\tag{G30}$$

Note that

$$(\Delta R)^2=2(\cosh^2 l_0-1)\frac{\theta_1\bar\theta_1\theta_2\bar\theta_2}{y_1y_2}=2(\cosh^2 l-1)\frac{\theta_1\bar\theta_1\theta_2\bar\theta_2}{y_1y_2},\quad \Delta R\cdot r=0.\tag{G31}$$

Finally, we will sum up the Grassmann even (G13) and odd (G29) contributions. We get

$$K(q_1,q_2|\tau)=-\frac{2ae^{-((1-a)/(2a))^2\tau}}{\pi\sqrt{2\pi\tau}}\int_{l_0}^{\infty}db\,\frac{1}{(\cosh b-\cosh l_0)^{1/2}}\Bigg[e^{-b^2/4\tau}\sinh\frac{b}{2a}$$
$$+\frac{\Delta R}{2}\frac{d}{db}\left(e^{-b^2/4\tau}\frac{\sinh(b/2a)}{\sinh b}\right)+\frac{(\Delta R)^2}{8}\frac{d}{db}\left\{\frac{1}{\sinh b}\frac{d}{db}\left(e^{-b^2/4\tau}\frac{\sinh(b/2a)}{\sinh b}\right)\right\}$$
$$+r\left\{(\cosh l-1)\frac{d}{db}\left(e^{-b^2/4\tau}\frac{\sinh(b/2a)}{\sinh b}\right)+e^{-b^2/4\tau}\left(\frac{b}{2\tau}\cosh\frac{b}{2a}+\frac{a-1}{2a}\sinh\frac{b}{2a}\right)\right\}\Bigg],\tag{G32}$$

We will change the region of the integration from (l_0,∞) to (l,∞) according to the relation, which holds for a nonsingular function $f(b)$ with $f(b)e^{(\mathrm{const})\cdot b}\to o(1)$, $b\to\infty$,

$$\int_{l_0}^{\infty} db \frac{f(b)}{(\cosh b - \cosh l_0)^{1/2}} = \int_{l}^{\infty} db \frac{f(b) - \frac{\Delta R}{2}\frac{d}{db}\left(\frac{f(b)}{\sinh b}\right) + \frac{(\Delta R)^2}{8}\frac{d}{db}\left\{\frac{1}{\sinh b}\frac{d}{db}\left(\frac{f(b)}{\sinh b}\right)\right\}}{(\cosh b - \cosh l)^{1/2}}, \tag{G33}$$

Proof: The lhs of the above equation is rewritten by

$$\int_{l_0}^{\infty} db \frac{f(b)}{(\cosh b - \cosh l_0)^{1/2}} = \int_{l}^{\infty} db \frac{f(b)}{(\cosh b - \cosh l_0)^{1/2}} + \int_{l_0}^{l} db \frac{f(b)}{(\cosh b - \cosh l_0)^{1/2}}. \tag{G34}$$

The first term is calculated as

$$\begin{aligned}(\text{The first term}) &= \lim_{\epsilon\to 0}\int_{l+\epsilon}^{\infty} db \left\{\frac{f(b)}{(\cosh b - \cosh l)^{1/2}} - \frac{\Delta R}{4}\frac{f(b)}{(\cosh b - \cosh l)^{3/2}} + \frac{3(\Delta R)^2}{32}\frac{f(b)}{(\cosh b - \cosh l)^{5/2}}\right\}\\ &= [\text{the rhs of (G33)}] - \mathscr{R},\end{aligned} \tag{G35}$$

where

$$\mathscr{R} \equiv \lim_{\epsilon\to 0}\left\{\frac{(\Delta R/2)(f(b)\sinh b) - \{(\Delta R)^2/8\}(1/\sinh b)(d/db)(f(b)/\sinh b)}{(\cosh b - \cosh l)^{1/2}} - \frac{[(\Delta R)^2/16][f(b)/\sinh b]}{(\cosh b - \cosh l)^{3/2}}\right\}\Bigg|_{b=l+\epsilon}. \tag{G36}$$

The second term can be evaluated as follows. Since

$$l - l_0 = (\Delta R/2)(1/\sinh l) + [(\Delta R)^2/8]\frac{\cosh l}{\sinh^3 l}, \tag{G37}$$

we find

$$\begin{aligned}(\text{The second term}) &= \lim_{\epsilon\to 0}\left[\frac{(l-l_0)f(b)}{(\cosh b - \cosh l_0)^{1/2}} - \frac{(l-l_0)^2}{2}\frac{d}{db}\left\{\frac{f(b)}{(\cosh b - \cosh l_0)^{1/2}}\right\}\right]\Bigg|_{b=l+\epsilon}\\ &= \mathscr{R} + \lim_{\epsilon\to 0}[\{\cosh(l+\epsilon) - \cosh l\}^{1/2}\times(\text{regular term})]\\ &= \mathscr{R}.\end{aligned} \tag{G38}$$

Summing up (G36) and (G38), we get the rhs of (G33). Q.E.D

Applying the formula (G33) to (G32), we get the kernel function (128).

APPENDIX H: HEAT KERNEL EXPANSION

In this appendix we show the following statements:

(1) $K(q,q|t) \sim \sum_{n=0}^{\infty} B_n(q,\Delta_{\rm SLB})t^{n/2}, \quad t\to 0^+$.

(2) $B_n(q,\Delta_{\rm SLB}) = 0$ for odd n, and for even n, B_n can be computed inductively and the result coincides with those in (149) and (150).

We will apply the method of the heat kernel expansion developed by Gilkey[48] and give some notations here. Let pairs (q^A,k_A) be the coordinates of the cotangent bundle $T^*(\mathrm{SH})$ and let α be a multi-index representing a set of non-negative integers, $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$. Let $|\alpha|$ and $\alpha!$ be defined by

$$|\alpha| \equiv \sum_{n=1}^{4}\alpha_n, \quad \alpha! \equiv \prod_{n=1}^{4}\alpha_n!. \tag{H1}$$

The derivative with respect to a^A and k_A are defined respectively by

$$\partial_q^\alpha \equiv (-i)^{|\alpha|}\left(\frac{\partial}{\partial q^{\bar\theta}}\right)^{\alpha_4}\left(\frac{\partial}{\partial q^{\theta}}\right)^{\alpha_3}\left(\frac{\partial}{\partial q^{\bar z}}\right)^{\alpha_2}\left(\frac{\partial}{\partial q^{z}}\right)^{\alpha_1},$$
$$\overleftarrow{\partial}_k^\alpha \equiv \left(\overleftarrow{\frac{\partial}{\partial k_z}}\right)^{\alpha_1}\left(\overleftarrow{\frac{\partial}{\partial k_{\bar z}}}\right)^{\alpha_2}\left(\overleftarrow{\frac{\partial}{k_\theta}}\right)^{\alpha_3}\left(\overleftarrow{\frac{\partial}{\partial k_{\bar\theta}}}\right)^{\alpha_4}. \tag{H2}$$

Note that we define ∂_q^α as a left derivative as usual, however, $\overleftarrow{\partial}_k^\alpha$ as a right derivative. The Fourier transformation of a "nice" function on SH is given by

$$\hat f(k) = \int d^4q\, e^{-iqk} f(q),$$
$$f(q) = \int d^4k\, e^{iqk}\hat f(k), \tag{H3}$$

where

$$qk \equiv q^A k_A. \tag{H4}$$

We find immediately the following formula:

$$\widehat{(\partial_q^\alpha)f(k)} = k^\alpha \hat f(k), \quad k^\alpha \equiv (k_{\bar\theta})^{\alpha_4}(k_\theta)^{\alpha_3}(k_{\bar z})^{\alpha_2}(k_z)^{\alpha_1}. \tag{H5}$$

A linear differential operator $P(q,\partial)$ of order m is

$$P(q,\partial) = \sum_{|\alpha|<m} a_\alpha(q)\partial_q^\alpha, \tag{H6}$$

where the $a_\alpha(q)$'s are certain functions of q. The symbol of the operator $P(q,\partial)$, denoted by $\sigma(P)$, is defined by

$$\sigma(P) = \sum_{|\alpha|\leqslant m} a_\alpha(q)k^\alpha, \tag{H7}$$

which is a polynomial in k. Due to (H5), the operation $P(q,\partial)$ on $f(q)$ is given by

$$(Pf)(q) = \int d^4k\,\sigma(P)\hat{f}(k). \tag{H8}$$

We can easily deduce the following formula for two linear differential operators P and Q,

$$\sigma(PQ) = \sum_\alpha (1/\alpha!)\,\{\sigma(P)\overleftarrow{\partial}{}_k^\alpha\}\{\partial_q^\alpha\sigma(Q)\}. \tag{H9}$$

We will apply the formula (H8) and (H9) to a wider class of functions $\sigma(P)$ than polynomials, and hence the corresponding operator P may not be a differential operator in general. Our method will be justified by the arguments in Ref. 48.

The super Laplace–Beltrami operator $\Delta_{\rm SLB}$ in (96) can be written by

$$\Delta_{\rm SLB} = -g^{AB}(q)\frac{\partial}{\partial q^B}\frac{\partial}{\partial q^A}, \tag{H10}$$

and hence its symbol becomes

$$\sigma(\Delta_{\rm SLB}) = g^{AB}(q)k_Bk_A. \tag{H11}$$

Using (H8) we find

$$\begin{aligned}(e^{-t\Delta_{\rm SLB}}f)(q) &= \int d^4k\,e^{iqk}\sigma(e^{-t\Delta_{\rm SLB}})\hat{f}(k)\\ &= \int\int d^4k\,d^4q'\,e^{i(q-q')k}\sigma(e^{-t\Delta_{\rm SLB}})f(q').\end{aligned} \tag{H12}$$

So the heat kernel associated with $\Delta_{\rm SLB}$ is given by

$$K(q,q'|t) = \int d^4k\,e^{i(q-q')k}\sigma(e^{-t\Delta_{\rm SLB}}). \tag{H13}$$

Our task has been reduced to calculating $\sigma(e^{-t\Delta_{\rm SLB}})$. According to the Cauchy's integral formula, a holomorphic function $f(z)$ is represented as

$$f(z) = \frac{1}{2\pi i}\oint_C d\lambda\,\frac{f(\lambda)}{z-\lambda}, \tag{H14}$$

where the direction of the integration is clockwise around z. For the Laplace–Beltrami operator, we define

$$e^{-t\Delta_{\rm SLB}} = \frac{1}{2\pi i}\oint_C d\lambda\,\frac{e^{-t\lambda}}{\Delta_{\rm SLB}-\lambda}, \tag{H15}$$

which leads to

$$\sigma(e^{-t\Delta_{\rm SLB}}) = 1/(2\pi i)\oint_C d\lambda\,e^{-t\lambda}\sigma(B_\lambda), \tag{H16}$$

where B_λ is the inverse operator to $\Delta_{\rm SLB}-\lambda$. Here, $\sigma(B_\lambda)$ will be expanded in some series of meromorphic functions of k;

$$\sigma(B_\lambda) = \beta_0 + \beta_1 + \cdots. \tag{H17}$$

Here each β_j is a sum of meromorphic functions of k and the order, which is defined as "(degree of numerator) − (degree of denominator)", of each of the meromorphic functions is $-(2+j)$. The β_j's are determined inductively as is shown below. Due to (H9) we have

$$\begin{aligned}1 &= \sigma(B_\lambda\cdot(\Delta_{\rm SLB}-\lambda))\\ &\sim \sum_{n=0}^\infty \sum_{|\alpha|+j=n}\frac{1}{\alpha!}[(\beta_j\overleftarrow{\partial}{}_k^\alpha)\{\partial_q^\alpha(\sigma(\Delta_{\rm SLB})-\lambda)\}].\end{aligned} \tag{H18}$$

Then we deduce that

$$\begin{aligned}\beta_0 &= \{g^{AB}(q)k_Bk_A-\lambda\}^{-1},\\ \beta_n &= -\beta_0\sum_{\substack{|\alpha|+j=n\\|\alpha|>0}}(\beta_j\overleftarrow{\partial}{}_k^\alpha)(\partial_q^\alpha g^{AB}k_Bk_A)\quad(n\geqslant 1).\end{aligned} \tag{H19}$$

We see that each β_n is expanded in terms of β_0. Therefore we write β_n such as

$$\beta_n = \sum_{j\leqslant 2n+1}\beta_{n,j}(q,k)(\beta_0)^j, \tag{H20}$$

where each $\beta_{n,j}(q,k)$ is a homogeneous polynomial of degree $(2j-n-2)$ in k. Thus we obtain

$$\begin{aligned}\sigma(e^{-t\Delta_{\rm SLB}}) &\sim \frac{1}{2\pi i}\oint_C d\lambda\Big(e^{-t\lambda}\sum_n\beta_n\Big)\\ &= \sum_{nj}\beta_{n,j}(q,k)\frac{1}{2\pi i}\oint_C d\lambda e^{-t\lambda}(g^{AB}k_Bk_A-\lambda)^{-j}\\ &= \sum_{nj}\frac{1}{(j-1)!}\beta_{n,j}(q,k)t^{j-1}e^{-t(g^{AB}k_Bk_A)}.\end{aligned} \tag{H21}$$

Equations (H13) and (H21) yield

$$\begin{aligned}&K(q,q|t)\\ &\sim \sum_{n=0}^\infty\sum_j\int d^4k\frac{1}{(j-1)!}\beta_{n,j}(q,k)t^{j-1}e^{-t(g^{AB}k_Bk_A)}.\end{aligned} \tag{H22}$$

Let us change the coordinates k for $\xi = t^{1/2}k$. Then,

$$\beta_{n,j}(q,k) = t^{n/2+1-j}\beta_{n,j}(q,\xi), \tag{H23}$$

so that

$$K(q,q|t) \sim \sum_{n=0}^\infty B_n(q,\Delta_{\rm SLB})t^{n/2}, \tag{H24}$$

where

$$B_n(q,\Delta_{\rm SLB}) = \sum_j\int d^4\xi\frac{1}{(j-1)!}\beta_{n,j}(q,\xi)e^{-g^{AB}\xi_B\xi_A}. \tag{H25}$$

Explicit calculation gives

$$\begin{aligned}B_0 &= -g^{1/2}/\pi,\\ B_2 &= -(g^{1/2}/\pi)(R/6),\\ B_4 &= -(g^{1/2}/\pi)\frac{1}{360}(5R^2-2R_{AB}R^{BA}+2R_{ABCD}R^{DCBA}),\\ &\ \ \vdots\\ B_1 &= B_3 = \cdots = 0.\end{aligned} \tag{H26}$$

[1] J. Hadamard, J. Math. Pure Appl. **4**, 27 (1898).
[2] For a recent review, see N. L. Balazs and A. Voros, Phys. Rep. **143**, 109 (1986).
[3] A. Selberg, J. Indian Math. Soc. **20**, 47 (1956).
[4] D. A. Hejhal, *Lecture Note in Mathematics*, Vol. 548 (Springer, New York, 1976).
[5] E. D'Hoker and D. H. Phong, Nucl. Phys. B **269**, 205 (1986); Phys. Rev. Lett. **56**, 912 (1986).
[6] D. Friedan, in *The Proceedings of the Workshop on Unified Theories*, edited by D. Gross and M. Green (World Scientific, Singapore, 1986).
[7] S. Matsumoto and Y. Yasui, Prog. Theor. Phys. **79**, 1022 (1988).
[8] S. Uehara and Y. Yasui, Phys. Lett. B **202**, 530 (1988).
[9] S. Uehara and Y. Yasui, J. Math. Phys. **29**, 2486 (1988).
[10] S. Matsumoto, S. Uehara, and Y. Yasui, Phys. Lett. A **134**, 81 (1988).
[11] M. A. Baranov, I. V. Frolov, and A. S. Schwarz, Teor. Mat. Fiz. **70**, 92 (1987).
[12] L. Crane and J. M. Rabin, Comm. Math. Phys. **113**, 601 (1988).
[13] Note that as to complex conjugation, we adopt such a convention that

$$\overline{X+Y} = \bar{X} + \bar{Y}, \quad \overline{XY} = \overline{YX}.$$

[14] Note that k is hyperbolic and hence $|a+d| > 2$.
[15] K. Aoki, Comm. Math. Phys. **117**, 405 (1988).
[16] P. Howe, J. Phys. A **12**, 393 (1979).
[17] In Ref. 10, the $(z,\bar{\theta})$-component of the metric g_{AB} is presented with a wrong sign.
[18] R. Casalbuoni, Nuovo Cimento A **33**, 389 (1976).
[19] The sign factor means that

$$(-)^A X_A = \begin{cases} X_A, & (A = z, \bar{z}), \\ -X_A, & (A = \theta, \bar{\theta}), \end{cases}$$

$$(-)^F = \begin{cases} 1, & F\text{:Grassmann even}, \\ -1, & F\text{:Grassmann odd}. \end{cases}$$

[20] The second solution is in fact obtained by taking a proper limit of the first solution (see Appendix C).
[21] A. M. Baranov, Yu. I. Manin, I. V. Frolov, and A. S. Schwarz, Comm. Math. Phys. **111**, 373 (1987).
[22] The solutions given in Ref. 10 are restricted ones.
[23] The norm of SPL (2,**R**) is a map onto $[1,\infty)$,

$$N{:}k\in \mathrm{SPL}(2,\mathbf{R}) \to N_k \in [1,\infty),$$

where N_k is defined by the maximum value of the eigenvalues of A_k. This norm is identical to the one induced by the metric on SH, as is seen below.
[24] We may interchange (u,μ) and (v,ν) in the following equations, which corresponds to changing the sign of ω, or the direction of motion.
[25] The last two conditions $\xi_1 = \xi_3 = 0$ are not necessary for $Z^{(I)}(t)$ to satisfy the condition (75) if $a = 2$ and $n > 0$, however, they are necessary to satisfy the condition (74).
[26] The explicit form of $H^{(2)}$ is

$$H^{(2)} = Y(\theta\bar{\theta}p_z p_{\bar{z}} - \theta p_z p_{\bar{\theta}} - \bar{\theta} p_{\bar{z}} p_\theta - p_\theta p_{\bar{\theta}}).$$

[27] F. A. Berezin, *Introduction to Superanalysis*, Math. Phys. and Appl. Math. Vol. 9, (Reidel, New York, 1987).
[28] D. V. Anosov, Proc. Steklov Inst. Math. **90** (1967).
[29] Ya. G. Sinai, *Introduction to Ergodic Theory* (Princeton U.P., Princeton, NJ, 1976).
[30] Ya. V. Pesin, Dokl. Akad. Nauk SSSR **226** (1976) [Sov. Math. Dolk. **17**, 196 (1976)].
[31] E. A. Margulis, Funkts. Anal. Ego Pril. **3**, 89 (1969).
[32] R. Browen, J. Math. **94**, 1 (1972).
[33] W. Parry and M. Pollicot, Ann. Math. **118**, 573 (1983).
[34] H. Huber, Math. Annalen **138**, 1 (1959).
[35] H. P. McKean, Comm. Pure Appl. Math. **25**, 225 (1972).
[36] M. Omote and H. Sato, Prog. Theor. Phys. **47**, 1367 (1972).
[37] T. Kawai, Found. Phys. **5**, 143 (1975).
[38] When $k = 0$, $A_\Lambda \sim y^\lambda$ or $y^{\lambda+1}$. Then A_Λ is unnormalizable and hence $k = 0$ will be excluded.
[39] We omit the suffix c here and hereafter for simplicity. Notice, however, that the supertrace of the $K_{\rm SRS}$ is independent of the parameter c, as will be seen below.
[40] S. Minakshisundaram and A. Pleijel, Can. J. Math. **1**, 242 (1949).
[41] S. Wolpert, Ann. Math. **115**, 501 (1982); **117**, 207 (1983).
[42] S. Uehara and Y. Yasui, Phys. Lett. B **217**, 479 (1989).
[43] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, New York, 1980).
[44] C. Grosche and F. Steiner, Phys. Lett. A **123**, 319 (1987).
[45] There is a misprint in the original formula presented in Ref. 43.
[46] S. Hitotsumatsu, S. Moriguchi, and K. Udagawa, *Suugaku–Koushiki* (Iwanami, Tokyo, 1975), Vol. III.
[47] We should examine the cases $0 < c \leqslant \frac{1}{2}$ and $-\frac{1}{2} \leqslant c < 0$, separately, however, the following result is valid when $|c| < \frac{1}{2}$.
[48] P. B. Gilkey, *The Index Theorem and the Heat Equation*, Math. Lect. Series (Publish and Perish, Wilmington, DE, 1956), Vol. 4.

Magnetic monopoles without string in the Kähler–Clifford algebra bundle: A geometrical interpretation

Adolfo Maia, Jr.
Department of Applied Mathematics, Universidade Estadual de Campinas, 13083 Campinas, SP, Brazil

Erasmo Recami
Department of Applied Mathematics, Universidade Estadual de Campinas, 13083 Campinas, SP, Brazil, and Dipartimento di Fisica, Università Statale di Catania, Catania, Italy, and I.N.F.N.-Sezione di Catania, Catania, Italy

Waldyr A. Rodrigues, Jr.
Department of Applied Mathematics, Universidade Estadual de Campinas, 13083 Campinas, SP, Brazil

Marcio A. F. Rosa
Department of Mathematics, Universidade Estadual de Campinas, 13083 Campinas, SP, Brazil

(Received 3 January 1989; accepted for publication 4 October 1989)

In substitution for Dirac monopoles with string (and for topological monopoles), "monopoles without string" have recently been introduced on the basis of a generalized potential, the sum of a vector A, and a pseudovector $\gamma_5 B$ potential. By making recourse to the Clifford bundle $\mathscr{C}(\tau M,g)$ $[(T_xM,g) = \mathbb{R}^{1,3}; \mathscr{C}(T_xM,g) = \mathbb{R}_{1,3}]$, which just allows adding together for each $x \in M$ tensors of different ranks, in a previous paper a Lagrangian and Hamiltonian formalism was constructed for interacting monopoles and charges that can be regarded as satisfactory from various points of view. In the present article, after having "completed" the formalism, a purely *geometrical interpretation* of it is put forth within the Kähler–Clifford bundle $\mathscr{K}(\tau^*M,\hat{g})$ of differential forms, essential ingredients being a generalized curvature and the Hodge decomposition theorem. Thus the way is paved for the extension of our "monopoles without string" to non-Abelian gauge groups. The analogy with supersymmetric theories is apparent.

I. INTRODUCTION

It is well known that, when describing the electromagnetic field $F_{\mu\nu}$ produced by a Dirac monopole[1] in terms of one single potential A_μ only, such a potential has to be singular along an arbitrary line starting from the monopole and going to infinity. This "string" has been considered—for a long time[2]—as *unphysical,* because the singularity in A_μ does not correspond to any singularity in $F_{\mu\nu}$.

It is also well known that, in the U(1) gauge theory of electromagnetism, which has as a mathematical model a Principal Fiber Bundle (PFB) $\pi: P \to M$ with group U(1), monopoles appear only if we consider a nontrivial bundle. Here, M is, in general, a four-dimensional Lorentzian manifold modeling the space time. The standard model is obtained by taking $M = \mathbb{R}^{1,3}$ and deleting from $\mathbb{R}^{1,3}$ the world line of the monopole. We then have as a model the PFB $\pi: P \to \mathbb{R}^2 \times S^2$ with group U(1) and the monopole charges appear as the Chern-numbers characterizing the PFB. These observations show that the topological theory does not put on equal footing the electric charge and the monopole, since the former is introduced through the electric current and the latter is a hole moving in space-time.[3,4] Notice that the topology of space-time becomes even more exotic when generalized monopoles are present.[5]

A way out has been looked for by many authors[2,6] via the introduction of a second potential B_μ. They did not completely succeed in dispensing with an exotic space-time whenever they wanted to stick to ordinary vector–tensor algebra. However, just on the basis of both a vector potential $A \in \sec \Lambda^1 \tau M \subset \sec \mathscr{C}(\tau M,g)$ [where $\mathscr{C}(\tau M,g)$ is the Clifford bundle constructed in the tangent bundle τM of the Lorentz manifold M equipped with the Lorentz metric g, and sec means a section of the bundle] and a pseudovector potential $\gamma_5 B \in \sec \mathscr{C}(\tau M,g)$, we recently constructed[7] a rather satisfactory formalism for magnetic monopoles without strings (i.e., living in the ordinary Minkowski space-time, $\mathbb{R}^{1,3}$), by making recourse to the Clifford algebra $\mathbb{R}_{1,3}$ or more precisely to the Clifford bundle $\mathscr{C}(\tau M,g)$ [where $(T_xM,g) = \mathbb{R}^{1,3}$]. Here, $\mathbb{R}_{1,3}$ is an algebra sufficiently powerful to allow adding together tensors of different ranks (grades). In Ref. 8, for example, both the electric and the magnetic current are vectorial, while in our approach they are represented by a vectorial and a pseudovectorial current, respectively (and nevertheless we can add them together[7]). Our formalism can be considered satisfactory for the reasons we shall see below. See also Ref. 9. Some analogous, but nonequivalent, results have been obtained in Refs. 10,11.

II. FROM CLIFFORD TO KÄHLER

In this paper we want, first of all, to pass from the $\mathscr{C}(\tau M,g)$ language, used in Ref. 7, to the $\mathscr{K}(\tau^*M,\hat{g})$ language, i.e., to the language of the differential forms in τ^*M, the cotangent bundle with metric $\hat{g}$ (equipped with the Kähler algebra).[12–14] This paves the way, incidentally, for a generalization of our "monopoles without string" to non-Abelian gauge groups.

The new language will allow us to approach the question of a suitable formalism for interacting charges and mono-

 0022-2488/90/020502-04$03.00

poles without string from a *geometrical* point of view in the space-time manifold.[15]

We recall that $\mathscr{K}(T_x^*M,\hat{g}) = \mathscr{C}(T_xM,g) = \mathbb{R}_{1,3}$, the so-called space-time algebra.[16] Now $\mathscr{K}(T_x^*M,\hat{g})$, as a linear space over the real field, can be written

$$\Lambda^0(T_x^*M) + \Lambda^1(T_x^*M) + \Lambda^2(T_x^*M) + \Lambda^3(T_x^*M) + \Lambda^4(T_x^*M), \quad (1)$$

where $\Lambda^k(T_x^*M)$ is the $\binom{4}{k}$-dimensional space of the k-forms. Here, $\Lambda(T_x^*M) = \Sigma\Lambda^k(T_x^*M)$ is called the Cartan algebra, and the pair $[\Lambda(T_x^*M),\hat{g}_x]$ is called the Hodge algebra. An analogous terminology exists for the vector bundles associated with these algebras.[9]

In $\mathscr{K}(\tau^*M,\hat{g})$ there is a particular differential operator ∂ odd in the $\mathbb{Z}_2$-gradation of the algebra.[17] To introduce ∂, consider first, for any $t^* \in \sec \tau^*M \subset \sec \mathscr{K}(\tau^*M,\hat{g})$ and any $t \in \sec \tau M$, the bilinear tensorial map of type (1,1) given by

$$\Psi \to t^*\nabla_t\Psi, \quad (2)$$

where Ψ is any element of $\sec \mathscr{K}(\tau^*M,\hat{g})$ and ∇_t is the covariant derivative of Ψ (considered as an element of the tensor bundle). Then ∂ is defined as the tensorial trace of the map:

$$\partial = \mathrm{Tr}(t^*\nabla_t). \quad (3)$$

In terms of a local basis $\{\gamma^\mu\}$ of one-form fields and its dual basis $\{e_\mu\}$ of vector fields, we can write

$$\partial = \gamma^\mu\nabla_{e_\mu}. \quad (3')$$

In particular, taking any local neighborhood $U \subset M$ with a local basis $\{dx^\mu\}$, so that $\partial = \gamma^\mu\nabla_\mu$, we can show[9,13] that for any $\Psi \in \sec(\Lambda\tau^*M,\hat{g}) \subset \sec \mathscr{K}(\tau^*M,\hat{g})$:

$$\partial\Psi = dx^\mu \wedge (\nabla_\mu\Psi) + \partial_\mu \lrcorner (\nabla_\mu\Psi), \quad (4)$$

where $\lrcorner$ is the usual contraction operator of the theory of differential forms. We have

$$dx^\mu \wedge (\nabla_\mu\Psi) = d\Psi, \quad (5)$$

$$\partial_\mu \lrcorner (\nabla_\mu\Psi) = -\delta\Psi, \quad (6)$$

where d is the usual differential, and δ is the Hodge coderivative operator, here defined as

$$\delta\Psi_k = (-1)^k *^{-1} d * \Psi_k \quad (7)$$

where $*$ is the Hodge star operator and $\Psi_k \in \sec \mathscr{K}(\tau^*M,\hat{g})$. The power of the Kähler bundle formalism appears clearly, once we add to the fundamental formula

$$\partial\Psi = (d - \delta)\Psi \quad (8)$$

the result[9,13,18,19]

$$\gamma^5\Psi_k = (-1)^t * \Psi_k, \quad (9)$$

where $\gamma^5 = \gamma^0\gamma^1\gamma^2\gamma^3$ is the volume element,[20] and where $t = 1$ for $k = 1,2,3$ and $t = 2$ for $k = 0,4$ in the particular case of the space time algebra $\mathbb{R}_{1,3}$ and with our conventions. We also have that $\partial^2 = (d - \delta)^2$ is the D'Alambertian operator.

III. GENERALIZED POTENTIAL AND FIELD: A SATISFACTORY FORMALISM

Before going on, observe that the "completed" Maxwell equations, $\delta F = -J_e, dF = -*J_m$, where $F \in \sec(\Lambda^2\tau^*M,\hat{g}) \subset \sec \mathscr{K}(\tau^*M,\hat{g})$ is the electromagnetic field and $J_e, J_m \in \sec(\Lambda^1\tau^*M,\hat{g}) \subset \sec \mathscr{K}(\tau^*M,\hat{g})$ are, respectively, the electric and magnetic currents, can be written as[21]

$$\partial F = J_e - *J_m = J_e + \gamma^5 J_m \equiv \overline{J}. \quad (10)$$

With the introduction of the generalized potential[21] $\overline{A} \equiv A + \gamma^5 B$, where $A,B \in \sec(\Lambda^1\tau^*M,\hat{g}) \subset \sec \mathscr{K}(\tau^*M,\hat{g})$, we get $F = \partial\overline{A} = \partial \wedge A + \partial\cdot(\gamma^5 B)$, once we impose the Lorentz gauge $\partial \circ A = 0$.[22] Then we can write Eq. (10) as:

$$\partial^2 A = J_e, \quad \partial^2 B = J_m. \quad (11)$$

In our previous work[7] we wrote Eqs. (10) and (11) in $\mathscr{C}(\tau M,g)$ instead of $\mathscr{K}(\tau^*M,\hat{g})$. There we succeeded in introducing a noncoventional Lagrangian that yields the correct field equations when varied with respect to the generalized potential. Our approach, however, cannot overcome the "no-go theorems" by Rosenbaum *et al.*[8]; for instance Rohrlich[8] showed that a single Lagrangian can yield both the field equations and the charge and pole motion equations *only* in the trivial case when $J_m = kJ_e$, where k is a constant. *Nevertheless* in our approach we need to apply the variational principle just once, since our Lagrangian[7] *implies* even the correct coupling of the currents to the field. In fact, as shown in detail in Refs. 9 and 23, the "completed" Maxwell equations [Eq. (10)] *imply*, if $S^\mu \equiv -\frac{1}{2}F\gamma_\mu F$, that

$$\partial_\mu S^\mu = F\cdot J_e + (\gamma^5 F)\cdot J_m, \quad (12)$$

where $S^\mu\gamma^\nu = E^{\mu\nu}$ is the symmetric energy-momentum of the electromagnetic field. Calling $K_e = F\cdot J_0$ and $K_m = -(\gamma^5 F)\cdot J_m$, and by projecting on the Pauli algebra $\mathbb{R}_{3,0}$, one does *consequently* find the expected expressions for the forces (in particular the *Lorentz forces*) acting on a charge and a monopole:

$$\mathbf{K}_e = \rho_e\mathbf{E} + \mathbf{J}_e \times \mathbf{H}, \quad (13a)$$

$$\mathbf{K}_m = -\rho_m\mathbf{E} + \mathbf{J}_m \times \mathbf{E}. \quad (13b)$$

IV. GENERALIZED CONNECTION AND CURVATURE

As is well known, in a gauge theory[24] the potentials are pullbacks of *connections* in the PFB $\pi:P\to M$ with group G, and the associated field is the pullback of the connection *curvature*. In the case of standard electromagnetism, the field $F \in \sec(\Lambda^2\tau^*M,\hat{g})$ is derived from a potential $A \in \sec(\Lambda^1\tau^*M,\hat{g})$, i.e.,

$$F = dA. \quad (14)$$

However the Hodge decomposition theorem[25] (valid for compact spaces) assures us that more generally, if $F \in \sec(\Lambda^2\tau^*M,\hat{g})$, then there always exist $A \in \sec(\Lambda^1\tau^*M,\hat{g})$, $*B \in \sec(\Lambda^3\tau^*M,\hat{g})$ and $C \in \sec(\Lambda^2\tau^*M,\hat{g})$, with $dC = \delta C = 0$, such that F can be *uniquely decomposed* into

$$F = dA + \delta * B + C. \quad (15)$$

The Hodge decomposition naturally suggests naming *generalized connection* the quantity

$$\bar{A} = A - *B \in \sec(\Lambda^1\tau^*M,\hat{g}) + \sec(\Lambda^3\tau^*M,\hat{g}) \quad (16)$$

and *generalized curvature*[26] the quantity

$$F = \partial\bar{A} = (d - \delta)\bar{A} = dA + \delta *B - d*B - \delta A. \quad (17)$$

Then

$$F \in \sec(\Lambda^0\tau^*M,\hat{g}) + \sec(\Lambda^2\tau^*M,\hat{g}) + \sec(\Lambda^4\tau^*M,\hat{g}).$$

If we want F to be still a two-form, then the last two addenda in Eq. (17) have to vanish, and we automatically end up with the Lorentz gauge condition

$$d*B = \delta A = 0, \quad (18)$$

and are left with

$$F = dA + \delta *B. \quad (17')$$

The field equations are obtained by evaluating ∂F, with $\partial \equiv d - \delta$:

$$(d - \delta)(dA + \delta *B - d*B - \delta A) = \partial^2 A - \partial^2 *B, \quad (19)$$

which writes

$$\partial F = J_e - *J_m \quad (20)$$

once we identify $\partial^2 A \equiv J_e$; $\partial^2 B \equiv J_m$. Equations (19) are of course the "completed" Maxwell equations, now deduced within a geometrical context via a natural generalization of the definitions of connection and curvature: a generalization inspired by the "correspondences" $\partial = d - \delta$ and $* = (-1)^r\gamma^5$, and by the Hodge decomposition theorem.

V. FURTHER REMARKS

(i) A rather interesting consequence of the geometrical interpretation just presented is that Eq. (17) can be assumed as a *new definition* of F, without imposing any longer the Lorentz gauge, since even in this case we get the right "completed" Maxwell equations [as it is clear from Eqs. (18) and (19)].

(ii) The introduction of our "monopoles without string" for the more general case of non-Abelian groups is discussed in Refs. 27 and 28. Here we want to emphasize once more that, for our aims, the ordinary tensorial language is too poor, since—among the others—it does not satisfactorily distinguish between scalar and pseudoscalar quantities, as on the contrary it is strictly required by physics. For instance, it is an essential character of the Lagrangian density of Ref. 7 to be the sum of a scalar and a pseudoscalar part.[7,29]

(iii) At last, let us take advantage of the present opportunity for pointing out some misprints that appeared in the previous paper,[7] that might make it difficult for the interested reader to rederive those results of ours: (1) at page 234, column 2, line 18: the two expressions $\partial\cdot\bar{J}$ ought rather to read $\partial\circ\bar{J}$; (2) at page 235, Eqs. (14) and (15): all three expressions should be written $\tilde{\bar{J}}\circ\bar{A}$; (3) at page 235: the last-term in the rhs of Eq. (17) ought to be eliminated; (4) at page 236, column 1, line 22: "pseudoscalars" should be corrected into "pseudovectors." Let us stress that the "ball product" $\circ$ is *not* a new fundamental product since in terms of the Clifford product we have, for $A, B \in \sec \mathscr{C}(\tau M,g)$, that $A\circ B \equiv \frac{1}{2}(A\tilde{B} + B\tilde{A})$.

ACKNOWLEDGMENTS

For stimulating, useful discussions, the authors are grateful to J. S. R. Chisholm, V. L. Figueiredo, E. Giannetto, A. Insolia, G. D. Maccarrone, F. Mercuri, R. Mignani, and particularly to E. Ferrari, M. Francaviglia, M. Novello, S. Ragusa, and A. Rigas. W. A. Rodrigues, Jr., is particularly grateful to the Dipart. Matematica, Univ. Trento, for kind hospitality.

This work is supported in part by INFN, CAPES, M.P.I., FAP-UNICAMP, C.S.F.N.eS.d.M., FAPESP. CNPq, CNR, and IBM-do-BRASIL.

[1]P. A. M. Dirac, Proc. R. Soc. London Ser. A **133**, 60 (1931); Phys. Rev. **74**, 817 (1948).

[2]N. Cabibbo and E. Ferrari, Nuovo Cimento **23**, 1147 (1964).

[3]Y. Choquet-Bruhat, C. de Witt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds and Physics* (North-Holland, Amsterdam, 1982), pp. 409–410. See also T. Eguchi, P. B. Gilkey, and A. J. Hanson, Phys. Rep. **66**, 213 (1980).

[4]C. N. Yang, N. Y. Acad. Sci. **294**, 86 (1977); E. Lubkin, Ann. Phys. **23**, 233 (1963).

[5]G. t'Hooft, Nucl. Phys. B **79**, 276 (1974); A. M. Polyakov, Sov. Phys. JETP Lett. **20**, 194 (1974).

[6]E. Ferrari, in *Tachyons, Monopoles and Related Topics*, edited by E. Recami (North-Holland, Amsterdam, 1978), p. 203 and references therein.

[7]M. A. Faria-Rosa, E. Recami, and W. A. Rodrigues, Jr., Phys. Lett. B **173**, 233 (1986); Erratum: **188**, 511 (1987).

[8]F. Rohrlich, Phys. Rev. **150**, 1104 (1966); D. Rosembaum, *ibid.* **147**, 891 (1966); E. F. Carter and H. A. Comen, Am. J. Phys. **41**, 994 (1973).

[9]W. A. Rodrigues, Jr., M. A. Faria-Rosa, A. Maia, Jr., and E. Recami, Report INFN/AE-88/2 (INFN, Frascati, 1988), to appear in Hadronic J. (1989).

[10]A. O. Barut, "Electromagnetic Fields on Manifolds: Betti numbers, Monopoles, and Strings, and Minimal Couplings," in *Differential Geometric Methods in Mathematical Physics*, edited by A. Dold and E. Eckmann (Springer, Berlin, 1978).

[11]J. S. R. Chisholm and R. S. Farwell, Proc. R. Soc. London Ser. A **337**, 1 (1981); Nuovo Cimento A **82**, 210 (1984).

[12]E. Kähler, Abh. Dtsch. Akad. Wiss. Berlin, Kl. Math. Phys. Tech. **1960**, No. 1 and No. 4.

[13]W. Graf, Ann. Inst. Henri Poincaré **29**, 85 (1978).

[14]Let us consider that the metric tensor $g \in \sec(\tau^*M \times \tau^*M)$ induces the "dual metric" $\hat{g}$ in the spaces $\Lambda^k(\tau^*M)$ (Ref. 3):

$$\hat{g}(\varphi_1,\varphi_2)\gamma^5 = \varphi_1 \wedge *\varphi_2,$$

where $\varphi_1,\varphi_2 \in \sec(\Lambda^k\tau^*M,\hat{g})$ is the so-called Hodge bundle. For future reference, note that, in the particular case in which $\varphi_1 = \varphi_2 = \varphi \in \sec(\Lambda^2\tau^*M,\hat{g})$, it holds:

$$\hat{g}(\varphi,\varphi) = -\hat{g}(*\varphi,*\varphi).$$

[15]For a *completely* geometric formulation (and generalization to arbitrary gauge groups) of the theory, we ought however to make recourse to a spliced bundle $\pi:P\circ P \to M$ with group $G \times G$ where M is an arbitrary space-time with nonzero, Lorentzian curvature: cf. Ref. 27.

[16]By adopting Hestenes' notations (cf. the second one of Ref. 13), we call *space-time algebra* the Clifford algebra $\mathbb{R}_{1,3}$ that we called "Dirac algebra" in Ref. 7. More correctly we shall reserve the name Dirac algebra for $\mathbb{R}_{4,1} \simeq \mathbb{C}(4)$. Notice, incidentally, that the *Majorana algebra* $\mathbb{R}_{3,1}$ is quite different from $\mathbb{R}_{1,3}$, so that *two* algebras [$\mathbb{R}_{1,3} \simeq \mathbb{H}(2)$, and $\mathbb{R}_{3,1} \simeq \mathbb{R}(4)$] can be naturally associated with Minkowski space-time; and this can have a bearing on physics (even for the mathematical problems with tachyons, for instance). At last, the *Pauli algebra* is $\mathbb{R}_{3,0} \simeq \mathbb{C}(2)$.

[17]Recall that we denote the Clifford product in $\mathscr{C}(\tau M,g)$, as well as in $\mathscr{K}(\tau^*M,\hat{g})$, by the mere *juxtaposition* of symbols.

[18]See, e.g., D. Hestenes and G. Sobczyk, *Clifford Algebra to Geometrical Calculus* (Reidel, Dordrecht, 1984); D. Hestenes, "A Unified Language for Mathematics and Physics" in *Clifford Algebras and their Applications in Mathematical Physics*, edited by J. S. R. Chisholm and A. K. Common (Reidel, Dordrecht, 1986); E. Tonti, Rend. Sem. Mat. Fis. Milano **46**, 163 (1976); P. Lounesto, Ann. Inst. Henri Poincaré A **33**, 53 (1980); I. Porteous, *Topological Geometry* (Cambridge U.P., Cambridge, 1981); W. A. Rodrigues, Jr., and V. L. Figueiredo, in *Proceedings of the 8th Italian Conference on General Relativity and Gravitational Physics*, edited by M. Cerdonio, R. Cianci, M. Francaviglia, and M. Toller (World Scientific, Singapore, 1989), pp. 467–471.

[19]N. Salingaros, J. Math. Phys. **22**, 1919 (1981).

[20]Recall that, whereas γ^5 is the volume element in $\mathscr{K}(\tau^*M,\hat{g})$, in Ref. 7 $\gamma_5 = e_0e_1e_2e_3 \in \mathscr{C}(\tau M,g)$ and $\{e_\mu\}$ is an orthonormal basis of $\mathbf{R}^{1,3}$.

[21]Cf. R. Mignani and E. Recami, Nuovo Cimento A **30**, 533 (1975) and references cited therein.

[22]Note that the scalar product between $\Psi_r \in \Lambda^r(T_x^*M)$ and $\Psi_k \in \Lambda^k(T_x^*M)$ is defined by $\Psi_r \cdot \Psi_k = \langle \Psi_r \Psi_k \rangle_{|r-s|}$; i.e., it is the component in $\Lambda^{|r-s|}(T_x^*M)$ of the Clifford product of Ψ_r and Ψ_k. Sometimes we make recourse also the *ball product* ($\circ$) which, in terms of the Clifford product, is defined as follows: $A \circ B = \frac{1}{2}(A\widetilde{B} + B\widetilde{A})$. The tilde operation, in its turn, is defined as follows: $D = d_1 d_2 \cdots d_r$; $\widetilde{D} = d_r \cdots d_2 d_1$, where the $d_i \in \mathbf{R}^{1,3}$, $i = 1,2,\ldots,r$.

[23]W. A. Rodrigues, Jr., E. Recami, A. Maia, Jr., and M. A. Faria-Rosa, Phys. Lett. **220**, 195 (1989).

[24]D. Bleecker, *Gauge Theory and Variational Principles* (Addison–Wesley, London, 1981).

[25]C. von Westenholz, *Differential Forms in Mathematical Physics* (North-Holland, Amsterdam, 1978), p. 321.

[26]Such a terminology is, of course, acceptable only when working in the base manifold. Despite this fact, the theory of electromagnetism with monopoles without string, containing the potentials A and B, can be formulated as a PFB $\pi: P \circ P \to M$ with group $U(1) \times U(1)$, where A and B are "parts" of a genuine connection in the sense of a PFB theory.[9,27]

[27]M. A. Faria-Rosa and W. A. Rodrigues, Jr., Mod. Phys. Lett. A **4**, 175 (1989).

[28]A. Maia, Jr., and W. A. Rodrigues, Jr., "Grassmann's fields and Generalized Magnetic Monopoles," Rev. Bras. Fís. **19**, 292 (1989).

[29]E. Recami, Riv. Nuovo Cimento **9**, issue no. 6 (1986): see pages 83 and 156 and references therein. Cf. also R. Mignani, E. Recami, and M. Baldo, Lett. Nuovo Cimento **11**, 586 (1972); E. Giannetto, *ibid.* **44**, 140, 145 (1985).

A turbulence model with stochastic soliton motion

S. Qian,[a] H. H. Chen, and Y. C. Lee
Department of Physics and Astronomy, The Laboratory for Plasma and Fusion Energy Studies, University of Maryland, College Park, Maryland 20742

(Received 19 May 1988; accepted for publication 20 September 1989)

A dissipative Benjamin–Ono equation is used to study fluid and plasma turbulence. The system is studied by an exact nonlinear mode truncation method in which a finite number of poles are used to present the solution. The justification of the pole expansion approach is discussed with the proof of a completeness theorem. The stability and spectrum analysis show that asymptotic behavior of the system is completely represented by a finite number of nonlinear modes. The behavior of those nonlinear modes resembles solitons, and exhibits a wide range of bifurcation phenomena and routes to turbulence.

I. INTRODUCTION

Solitons have been widely studied in the last decade since the discovery of the integrability of the Korteweg–de Vries (KdV) equation. Many nonlinear evolution equations have been solved by the inverse scattering method (IST), including the KdV equation, the nonlinear Schrödinger equation, etc.

A particular system of interest to us is the Benjamin–Ono (BO) equation[1,2]

$$\frac{\partial u}{\partial t} + 2u\frac{\partial u}{\partial x} + \mathscr{H}\frac{\partial^2 u}{\partial x^2} = 0, \tag{1}$$

where the Hilbert transform $\mathscr{H}$ is defined as

$$\mathscr{H} f(x) = \frac{1}{\pi}\mathscr{P}\int_{-\infty}^{\infty}\frac{f(x')}{x'-x}\,dx'.$$

It describes the internal waves in a fluid or plasma two-fluid system. Its Lax pairs have been constructed and the related inverse scattering transform for infinite boundary conditions has been solved.

It is interesting to ask whether the solitonlike behavior also exists in nonintegrable systems. Despite the power of IST methods for integrable equations, they are of little assistance in tackling problems related to nonintegrable systems, which are usually characterized by randomness and nonpredictability.

In this paper, we introduce such a system—the dissipative BO (DBO) equation,

$$\frac{\partial u}{\partial t} + \mu\mathscr{H}\frac{\partial u}{\partial x} + \beta\mathscr{H}\frac{\partial^2 u}{\partial x^2} - \nu\frac{\partial^2 u}{\partial x^2} + 2u\frac{\partial u}{\partial x} = 0. \tag{2}$$

It allows an exact truncation in nonlinear normal modes that behave like solitons in the integrable system. These nonlinear normal modes are localized pulses. They preserve their identities while interacting strongly with each other. However, their motion could be highly random and unpredictable.

The powerful IST method that deals with the integrable system is not effective for the current system. We apply instead the pole expansion method to represent the solution analytically. This method is presented in Sec. II. In Sec. III, the pole expansion solutions are analyzed. We study both the linear and nonlinear structural stabilities of the pole expansion solutions, and conclude that the number of poles in the asymptotic state is determined by growth and damping. Section IV starts with a discussion of the completeness of the pole representation. We found that a variant of the scattering equation derived from those of the BO equation could be especially useful in exhibiting the pole solutions of this nonintegrable equation. These analyses suggest that the pole representation is complete. A proof of the completeness for the pole expansion method is also presented in this section. Dynamics of the poles are examined in Sec. V. The soliton behavior, bifurcations of soliton orbits, and the chaotic motions of poles are discussed. We conclude with a summary in Sec. VI.

II. THE POLE EXPANSION METHOD

The idea of using the pole expansion method to analyze solutions of nonlinear wave equations was initiated by Kruskal. He used the pole representation to study the soliton solutions of the KdV equation.[3] Subsequently, Moser,[4] Airault *et al.*,[5] and the Choodnovsky brothers[6] studied it for other equations. Later, Chen and Lee applied it to solve the BO equation and the nonintegrable DBO equation.[7,8] It was shown that the solution of the BO equation can be expressed as a linear superposition of simple poles moving on the complex plane. With infinite boundary conditions, we have

$$u(x,t) = \sum_j \frac{-i}{x - x_j(t)} + \text{c.c.}$$

The poles are always complex conjugate pairs for real solutions. Later in the paper, only the poles in the upper half complex plane are mentioned. The solution with one pole is a permanent pulse traveling with uniform speed

$$u(x-vt) = 2v/[4v^2(x-vt)^2 + 1].$$

It is therefore a soliton.

A dissipative version of BO was introduced to describe turbulent flows in Ref. 8:

$$\frac{\partial u}{\partial t} + \mu\mathscr{H}\frac{\partial u}{\partial x} + \beta\mathscr{H}\frac{\partial^2 u}{\partial x^2} - \nu\frac{\partial^2 u}{\partial x^2} + 2u\frac{\partial u}{\partial x} = 0. \tag{2}$$

[a] Present address: Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545.

 0022-2488/90/020506-11$03.00

This equation embodies all the essential characteristics of fluid and plasma turbulence: namely, dispersion, nonlinearity, dissipation, and growth. What distinguishes this model from other turbulence models is that the present model can be solved exactly by pole expansion.

In the more realistic periodic boundary conditions, the pole expansion solution of Eq. (2) is

$$u(x,t) = -\frac{i\beta+\nu}{2}\sum_{j=1}^{n}\cot\left(\frac{x-x_j(t)}{2}\right) + \text{c.c.}, \tag{3}$$

where the period is normalized to 2π and the x_j move in the upper half plane. The cotangent function can be expanded as

$$\cot\left(\frac{x-y}{2}\right) = \sum_{l=-\infty}^{\infty}\frac{2}{x-y+2l\pi}.$$

It has a single pole in each period. To derive the dynamical equation for the poles, we Fourier transform Eq. (3) to obtain

$$a_k = (\beta+i\nu)\sum_{j=1}^{n} e^{-ikx_j^*}, \quad \text{for } k>1,$$

where

$$u(x,t) = \sum_{k=-\infty}^{\infty} a_k e^{ikx},$$

with $a_{-k} = a_k^*$. The equation for the k th Fourier component becomes

$$\sum_{j=1}^{n}\left\{\frac{dx_j^*}{dt} - i\mu + (-i\beta+\nu)\sum_{l\neq j}^{n}\cot\left(\frac{x_j^*-x_l^*}{2}\right) + (i\beta+\nu)\sum_{l=1}^{n}\cot\left(\frac{x_l-x_j^*}{2}\right)\right\}e^{-ikx_j^*} = 0.$$

Since this is true for all k, we arrive at the equation of poles

$$\frac{d}{dt}x_j(t) = -i\mu - (i\beta+\nu)\sum_{l\neq j}^{n}\cot\left(\frac{1}{2}(x_j-x_l)\right) + (i\beta-\nu)\sum_{l=1}^{n}\cot\left(\frac{1}{2}(x_j-x_l^*)\right). \tag{4}$$

In the infinite period limit, where

$$u(\pm\infty,t) = 0,$$

we have

$$u(x,t) = -(i\beta+\nu)\sum_{j=1}^{n}\frac{1}{x-x_j} + \text{c.c.} \tag{5}$$

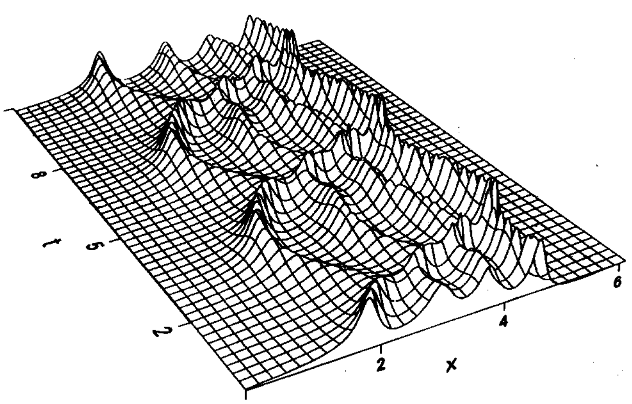


FIG. 1. A periodic solution of a four-pole with $\gamma_1 = 0.07$, $\gamma_2 = 0.18$.

and

$$\frac{dx_j}{dt} = -i\mu - (i\beta+\nu)\sum_{l\neq j}^{n}\frac{2}{x_j-x_l} + (i\beta-\nu)\sum_{l=1}^{n}\frac{2}{x_j-x_l^*}. \tag{6}$$

We note that Eqs. (5) and (6) approach the BO limit when μ and ν approach zero.

An interpretation of Eq. (3) is to consider each cotangent term with its complex conjugate as a nonlinear normal mode. The expansion resembles superposition in a linear system. These nonlinear modes are coupled with each other by Eq. (4). The nonlinear mode decompositions are exact, and assume great advantage over the usual linear mode decomposition method to study turbulence models since the latter involves an infinite degree of freedom while the former involve only a finite number of them.

The nonlinear modes are solitonlike wave pulses. They preserve their identities in the dynamics, very much like solitons in the ordinary integrable KdV equation. In Fig. 1 we present a numerical obtained four-pole periodic solution. However, solitons in this dissipative system are not the same as solitons in a conservative system. For instance, the DBO solitons may grow or damp by receiving energy from other solitons, and after a collision both speeds and amplitudes can change in contrast to the elastic collision between solitons in the BO equation. In Fig. 2 the collision of two solitons in the DBO equation is plotted.

III. ASYMPTOTIC STATE

If the motion of the poles is measured in a frame moving with velocity v_0, Eq. (4) becomes

$$\frac{dx_j(t)}{dt} = -i\mu - v_0 - (i\beta+\nu)\sum_{l\neq j}^{n}\cot\left(\frac{x_j-x_l}{2}\right) + (i\beta-\nu)\sum_{l=1}^{n}\cot\left(\frac{x_j-x_l^*}{2}\right). \tag{7}$$

The center of poles on the complex plane is defined as

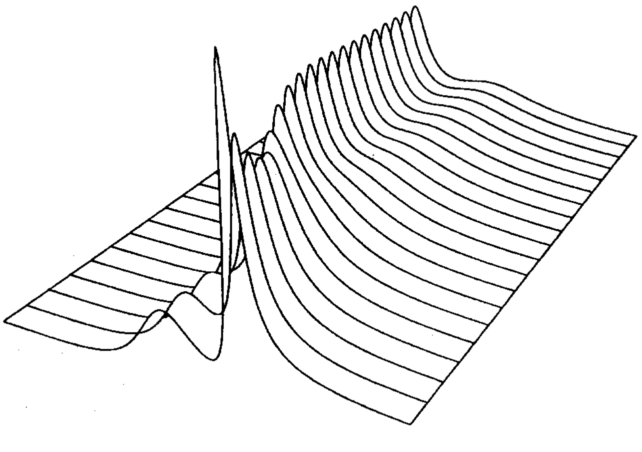


FIG. 2. Collision of two DBO solitons in a two-pole case. Both pulses change height during the collision. The pulse at right before the collision slows down and the other pulse gains speed during the collision.

$$X = \mathrm{Re}\, X + i\, \mathrm{Im}\, X = \sum_{j=1}^{n} x_j. \quad (8)$$

Summing over all indices j in Eq. (7), we have center velocity

$$\frac{dX}{dt} = -(i\mu + v_0)n + (i\beta - \nu)\sum_{j,l} \cot\left(\frac{x_j - x_l^*}{2}\right). \quad (9)$$

Choosing $v_0 = \mu\beta/\nu$, Eq. (9) gives

$$\mathrm{Re}\, X - (\beta/\nu)\mathrm{Im}\, X = \text{const.} \quad (10)$$

Therefore, in the frame that moves with velocity $v_0 = \mu\beta/\nu$, the center of poles moves on a straight line which makes an angle θ to the real axis:

$$\theta = \arctan(\nu/\beta).$$

When the center shifts on the real axis, it also shifts on the imaginary axis.

Equation (7) gives us some general idea about how poles move in the complex plane. First, no pole can move across the real axis. When a pole x_j moves close to the real axis in the upper half complex plane, its image, x_j^*, moves to the real axis in the lower half plane simultaneously. The interaction between them becomes very large and Eq. (7) becomes

$$\frac{dx_j}{dt} \simeq (i\beta - \nu)\cot\left(\frac{x_j - x_j^*}{2}\right) \simeq \frac{\beta + i\nu}{\mathrm{Im}\, x_j},$$

so that

$$\mathrm{Im}\, x_j(t) \simeq \sqrt{(\mathrm{Im}\, x_j(0))^2 + 2\nu t}.$$

When the imaginary part of x_j approaches zero, an infinitely larger repelling velocity prohibits any pole from moving across the real axis.

We now turn to the case when two poles collide. When x_1 is very close to x_2, Eq. (7) is approximated by

$$\frac{dx_1}{dt} \simeq -(i\beta + \nu)\cot\left(\frac{x_1 - x_2}{2}\right),$$

$$\frac{dx_2}{dt} \simeq -(i\beta + \nu)\cot\left(\frac{x_2 - x_1}{2}\right),$$

and

$$\frac{d}{dt}(x_1 - x_2) \simeq -\frac{4(i\beta + \nu)}{x_1 - x_2}. \quad (11)$$

Assume $x_{12} = x_1 - x_2$, we have the solution

$$x_{12}^2(t) = x_{12}^2(0) - 8(i\beta + \nu)t.$$

In the above approximation, the trajectory is not time reversible if the damping ν is nonzero, which shows that the trajectory x_{12} is not symmetric to the imaginary axis. As a consequence of the asymmetric trajectory, the pulse velocities and amplitudes are changed in the collision. That is obvious because energy is not conserved in the inelastic collision. In an elastic collision ($\nu = 0$), two solitons will resume their speeds and heights. That has been the case in the BO equation. Since collisions in the DBO equation are inelastic, both velocities of the solitons and their heights are changed, as shown in Fig. 2.

In the case of the n-soliton solution, the general behavior of solitons is analytically intractable. However, we can investigate their structural stability. The analysis reveals that only a finite number of poles survive in the time asymptotic limit. If the system is nonlinearly stable, i.e., all poles are confined in a finite region, then the time averaged center velocity should be zero. The imaginary part of the center velocity of poles has[9]

$$\frac{d}{dt}\mathrm{Im}\, X = n(2n\nu - \mu) + \frac{\nu}{4(\beta^2 + \nu^2)}\langle (u - \langle u \rangle)^2 \rangle,$$

where angle brackets indicate a spatial average. This gives us the following theorem.[9]

Theorem I: The n pole solution cannot be nonlinearly stable if

$$2n\nu - \mu > 0.$$

Assume a wave approximated by n poles is perturbated by a small perturbation represented by m poles with very large imaginary part, the center velocity δX of m poles has[9]

$$\frac{d}{dt}\mathrm{Im}\, \delta X \simeq 2m(2n + m)\nu - \mu. \quad (12)$$

The center δX will shift to imaginary infinity if

$$2m(2n + m)\nu - \mu > 0.$$

None of the m poles can come down to joint the finite configuration of an n pole, we then obtained[9] the following theorem.

Theorem II: The n pole solution is linearly stable, i.e., the pole at infinity would not move into the finite region, if

$$n \geqslant \mu/4\nu - \tfrac{1}{2}.$$

Combining Theorems I and II, we obtain the condition for the number of stable nonlinear modes,

$$\mu/2\nu > n \geqslant \mu/4\nu - \tfrac{1}{2}. \quad (13)$$

If we call μ/ν the Reynold's number, we see that the number of stable nonlinear modes is finite and linearly increases with the Reynold's number. Since Theorem II gives only the condition for a solution to be nonlinearly unstable, a nonlinearly stable solution requries a stronger condition. In the numerical simulations, a stable solution always prefers the smallest n ($n_{\min}$) which satisfies Eq. (13). If a solution is represented by n poles with $n > n_{\min}$, numerical solutions show that those extra poles will move to imaginary infinity and the related nonlinear modes will decay. On the other hand, a solution truncated at n with $n < n_{\min}$ is structurally unstable. If one more pole is added to the system, it will stay in the finite region. Therefore, an n-pole solution is stable if

$$\frac{1}{2(2n-1)} \geqslant \frac{\mu}{\nu} > \frac{1}{2(2n+1)}. \quad (14)$$

In the next section, we will show that dynamic behavior of a smooth initial wave function can be represented completely by the dynamics of poles. According to the above two stability theorems, the number of poles that will remain in the finite region is determined by the ratio of growth to damping. Therefore, in the asymptotic state, it is sufficient to study n-pole dynamics with $n = n_{\min}$.

IV. COMPLETENESS THEOREM

We have shown that pole expansion can be exactly truncated for the dissipative BO equation. However, it is important to know whether the pole representation is complete and converges to a solution of the DBO equation. We would like to study this problem before we go on to discuss the general properties of the solutions.

For a given periodic wave function u, we want to know under what conditions the function can be represented by pole expansion,

$$u(x,t) = \sum_j \psi_j(x,t), \tag{15}$$

where

$$\psi_j(x,t) = \eta\left\{\cot\left(\frac{x - x_j(t)}{2}\right) - i\right\} + \text{c.c.}, \tag{16}$$

x_j is in the upper half plane, and $\eta = -(i\beta + \nu)$ is a complex number. A constant has been subtracted from each mode so that $\psi_j \to 0$ as $x_j \to i\infty$ and the series could be convergent.

We start with the finite-pole approximation of u:

$$u_m(x,t) = \sum_{j=1}^{m} \psi_j(x,t). \tag{17}$$

Its Fourier components are

$$a_k^m = -i\eta^* \sum_{j=1}^{m} e^{-ikx_j^*}, \quad \text{if } k>0. \tag{18}$$

Suppose u has a uniformly convergent Fourier expansion,

$$u(x,t) = \sum_{k=-\infty}^{\infty} a_k e^{ikx}. \tag{19}$$

If $u_m(x,t)$ and u are identical, we must have

$$a_k = -i\eta^* \sum_{j=1}^{m} e^{-ikx_j^*}, \quad k>0. \tag{20}$$

The equations for negative k are determined by the complex conjugate of Eq. (20) because u is real. Equation (20) can be rewritten as

$$c_k = \sum_{j=1}^{m} w_j^k, \quad k = 1,2,...,\infty, \tag{21}$$

where $c_k = ia_k/\eta^*$ and $w_j = e^{-ix_j^*}$. Obviously, $0<|w_j|<1$, and w_j are located within the unit circle since the poles x_j are in the upper half plane. To determine the poles for a given u, we need to solve Eq. (21), which consists of an infinite number of polynomial equations. Approximately, we solve only the first set of n equations,

$$c_k = \sum_{j=1}^{m} w_j^k, \quad k = 1,2,...,n. \tag{22}$$

If we assume $m = n$, the w_j are uniquely determined by a corresponding nth-order polynomial,

$$P_n(w) = \prod_{j=1}^{n}(w + w_j) = \sum_{j=0}^{n} \alpha_j w^{n-j}, \tag{23}$$

where

$$\alpha_i = \sum_{n>j_1>j_2>\cdots>j_i>1} w_{j_1} w_{j_2} \cdots w_{j_i}. \tag{24}$$

To relate α_i and c_k, we use induction to find the recursion relation

$$\alpha_0 = 1, \quad \alpha_1 = c_1, \quad \alpha_{n+1} = \frac{1}{n+1} \sum_{i=1}^{n+1} (-1)^{i+1} \alpha_{n+1-i} c_i. \tag{25}$$

From Eq. (24), α_{n+i} equals zero if u has only n poles, because we have at most n different w_j to choose. We assume that the polynomial $P_n(w)$ does not have zero roots, because zero roots do not contribute to u in the pole expansion.

A new polynomial $Q_n(z)$ which has roots $z_i = 1/w_i$ is defined as

$$Q_n(z) = \sum_{i=0}^{n} \alpha_i z^i. \tag{26}$$

Now let n go to infinity; then

$$Q(z) = \sum_{i=0}^{\infty} \alpha_i z^i. \tag{27}$$

For the function with one Fourier component, $u = a_k e^{ikx} + \text{c.c.}$, $Q(z)$ can be solved easily:

$$Q(z) = \exp(-ia_k(-z)^k/\eta^* k). \tag{28}$$

Assuming $u = u_a + u_b$, the two power series $Q_a(z)$ and $Q_b(z)$ can be formulated separately from the Fourier coefficients of u_a and u_b. One can prove that the power series Q, for u, is exactly the product $Q_a(z)Q_b(z)$. Therefore, the power series $Q(z)$ has the general form

$$Q(z) = \exp(-R(z)), \tag{29}$$

where

$$R(z) = \frac{-\beta + i\nu}{\nu^2 - \beta^2} \sum_{l=1}^{\infty} \frac{a_l}{l}(-z)^l.$$

An n-pole approximation is obtained by solving for the zeros of the partial sum Q_n numerically. When n approaches infinity, poles may have no limit or all poles tend to imaginary infinity. As an example, considering the single mode case, the partial sum of Eq. (28) has n zeros located on the annulus[10]

$$0.2k\,\text{Int}(n/k)|\eta| < |a_k|\,|z|^k < (\text{Int}(n/k) + \tfrac{4}{3})k\,|\eta|,$$

where Int means integer part. When n approaches infinity, all poles approach infinity. Although each nonlinear normal mode contributes infinitesimally, their total contribution is finite. It is important to point out that the pole expansion is not unique. If $m>n$ in Eq. (22), we can have many m-pole approximations that exactly represent the first n Fourier modes. We have discussed a $m = 2n$ expansion when u is decomposed into two parts $(u = u_a + u_b)$. Similarly, an $(n\times n)$-pole expansion can be found by solving Eq. (28) for each Fourier mode. This leads to the following theorem.

Theorem III: If a function u, which is periodic in x, is bounded and piecewise smooth, then, for any ϵ at given t, there is an N such that, for any $n>N$,

$$|u_n(x,t) - u(x,t)| \leqslant \epsilon,$$

where $u_n(x,t)$ is the n-pole approximation of u.

A detailed proof is given in Appendix A.

Theorem III shows that the pole expansion is a complete representation, like the Fourier series, and that any piecewise smooth periodic function can be approximated by finite number of poles. A special propertity of the DBO equation is that the finite pole expansion is an exact solution and evolution of poles is simply obtained from Eq. (4). To show that pole expansion is complete for the solutions of the DBO equation, we need to prove that a small difference $\epsilon = u - u_n$ at $t = 0$ remains small for any $t > 0$, where u_n is an n-pole expansion. If the true solution u is assumed to be piecewise smooth, then ϵ is piecewise smooth and can be approximated by m poles y_j. In Theorem II we have shown that the center velocity of poles y_j will move to imaginary infinity when n is larger than or equal to $\mu/4\nu - 0.5$. Assuming the imaginary parts of y_j are very large, they satisfy the equations

$$\frac{d}{dt}\operatorname{Im} y_j = 2\nu(2n+1) - \mu - 2\operatorname{Im}\left\{(i\beta + \nu)\sum_{l \neq j}^{m} \frac{i\exp(iy_j - iy_l)}{\exp(iy_j - iy_l) - 1}\right\}.$$

As a result of the redundancy of the pole expansion, it is always possible to represent ϵ with poles near imaginary infinity. Whether a pole y_j moves to infinity depends on the interaction with all other poles. While the interaction with poles in the finite region (x_i) apparently forces y_j to move away from the real axis, the interaction with other poles in the perturbation could push y_j down to the real axis when some y_l are very close to y_j. However, when y_j is pushed down, it separates from the other pole and the interaction is weakened. Eventually, y_j will move to imaginary infinity again. We have tested many cases where a finite pole configuration is perturbated by m poles whose imaginary part is greater than any pole in the finite configuration. In all cases, poles in the perturbation move sooner or later to infinity, no matter how the poles are distributed. Therefore, it seems that ϵ will eventually decay when n is very large, because not only does the center of the m poles move to infinity, but so does every pole. We conjecture that the finite-pole expansion is a complete set of solutions.

We further examine the completeness by comparing the pole expansion solutions with the solutions integrated by other methods. The DBO equation can be numerically integrated by either the pole expansion or the split Fourier method.[11] If the solution is well behaved and the numerical scheme is stable, then the approximate solution of the split Fourier method can be arbitrarily close to the exact solution. On the other hand, if the pole expansion solution is not complete, one could expect solutions from the split Fourier method that do not agree with that obtained from the pole expansion.

We carried out a large number of numerical studies in which solutions of the pole expansion are compared with the solutions of the split Fourier method. Cases studied cover a wide region of parameters (damping, dispersion, and dissipation) and initial conditions. In Fig. 3, a pole expansion solution and a solution from the split Fourier method with the same initial condition are compared. The initial condition is randomly chosen and is approximated by 120 poles. In the numerical integrations, only poles with imaginary part less than 20 are kept. There are only ten poles left when the pole equation is integrated to $t = 6$ in this example. The $\mathscr{L}_2[0,2\pi]$ norm of the difference of the two solutions is

$$\epsilon = \int_0^{2\pi} |u_p(x,t) - u_f(x,t)|\,dx,$$

where u_p is the pole expansion solution and u_f is the solution from the split Fourier method, which is on the order of 10^{-5}, very small compared to the order unity $\mathscr{L}_2$ norm of u_p. Both u_p and u_f converge to the same asymptotic state when they are integrated for a longer period of time. In the region where the asymptotic states are stationary or periodic (Fig. 4), the

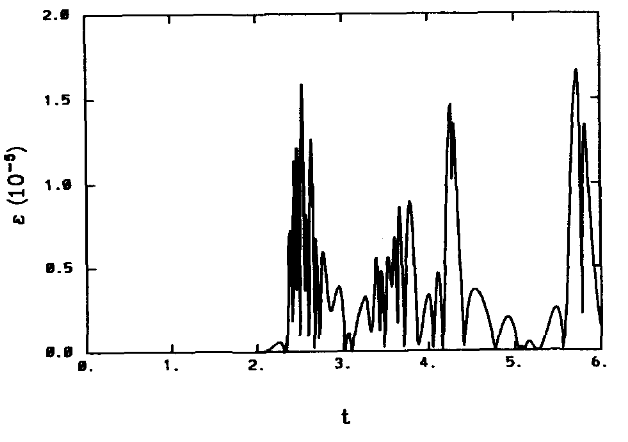


FIG. 3. Comparison of the pole expansion method and the split operator method. Both cases are integrated from the same initial condition with $\gamma_1 = 0.07$ and $\gamma_2 = 0.25$. In the pole expansion case, solution is represented by 120 poles. The solution in the split operator method is truncated at the 128th Fourier mode. The $\mathscr{L}_2$ norm of the difference between the two solutions from $t = 0$ to 6 is plotted in the figure. In this period, 110 poles move away from the real axis to the region where the imaginary part is greater than 20.

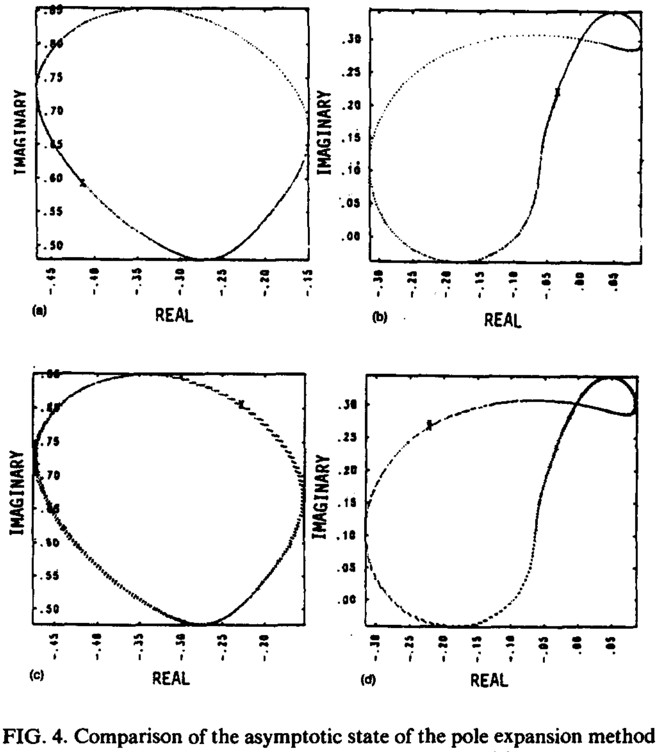


FIG. 4. Comparison of the asymptotic state of the pole expansion method and the split operator method in a periodic case with $\gamma_1 = 0.07$ and $\gamma_2 = 0.185$. (a) and (b) are trajectories of the first two Fourier coefficients from the pole expansion solution, which are exactly the same as (c) and (d) from the split Fourier method.

results show that both methods give identical asymptotic solutions. In the cases for which the asymptotic solutions are chaotic in the pole expansion method, the split Fourier method also obtains chaotic solutions, although it is difficult to determine if they are really identical. The numerical studies show that the solutions of the pole expansion and the solutions of the split Fourier method converge to the same solutions in all cases studied.

An initial m-pole wave will converge to an n-pole time-asymptotic solution ($m>n$) with n satisfying relation (6). All $m-n$ poles move rapidly to imaginary infinity and related nonlinear mode decay. How an initial wave evolves to the time-asymptotic solution in the split Fourier method can be studied in the scattering data analysis used to determined the number of poles (number of eigenvalues) of any wave. The scattering equation is derived from the scattering equation for the BO equation.[12] The Lax pair for the BO equation is the following:

$$i\frac{\partial\psi^+}{\partial x}+P^+u\psi^+-\frac{\lambda}{2}[\psi^+(i\infty)+\psi^-(-i\infty)]=-\lambda\psi^+, \tag{30a}$$

$$i\frac{\partial\psi^+}{\partial t}+\frac{\partial^2\psi^+}{\partial x^2}+2i\lambda\frac{\partial\psi^+}{\partial x}-2i\psi^+P^+\frac{\partial u}{\partial x}=0, \tag{30b}$$

where $+$ ($-$) means $\psi^\pm$ are analytic in the upper (lower) half plane, and

$$P^\pm=\tfrac{1}{2}(1\pm\mathscr{H})$$

is the projection operator. In the BO equation, the scattering data reveal many properties of the pole expansion solutions. For example, there is a connection between eigenvalues and the poles in the solution u. Here we apply the first scattering equation, Eq. (30a), to the DBO equation to find the number of poles from the eigenvalues.

When u is an n-pole expansion, Eq. (30a) with a periodic boundary condition can be reduced to an equation of $n\times n$ matrices

$$|\mathscr{M}+\lambda\mathscr{I}|=0,$$

where $\mathscr{I}$ is the unit matrix and the matrix elements of $\mathscr{M}$ are given in Appendix B. Hence the number of eigenvalues in Eq. (30a) equals the number of poles in the function u. Using the pole equation of the DBO equation, we can calculate the time dependence of eigenvalues.

Numerical calculation of the scattering equation for any function is more conveniently done in Fourier space. The scattering equation (30a) with the same boundary conditions is transferred to

$$-ka_k+\lambda a_k+\sum_{l=1}^{\infty}u_{k-l}a_l=0,\quad k=1,2,\ldots,\infty, \tag{31a}$$

$$\sum_{l=1}^{\infty}u_{-l}a_l=\frac{\lambda}{2}[\psi^+(i\infty)+\psi^-(-i\infty)], \tag{31b}$$

where $a_0=0$ and

$$\psi^+(x)=\sum_{k=1}^{\infty}a_ke^{ikx},$$

$$u(x)=\sum_{l=-\infty}^{\infty}u_le^{ilx}.$$

Note that Eq. (31b) merely gives the relation between ψ^+ and ψ^-, Eq. (31a) is sufficient to solve for the eigenvalues λ's. Since Eqs. (31a) and (30a) are solved in the upper half plane with the same boundary values, the two equations must have exactly the same eigenvalues. However, when Eq. (31a) is truncated at $k_{\max}$ and solved numerically, in addition to n eigenvalues very close to those solved in Eq. (30a), we also obtain many eigenvalues close to integer $n+1$, $n+2,\ldots$, because of truncation. Therefore, the number of noninteger eigenvalues in Eq. (31a) equals the number of poles when u has the pole expansion.

Numerically, we integrate the wave function u by the split Fourier method, and use that to determine the time evolution of the eigenvalues. The time evolution of λ's are plotted in Fig. 5. Figure 5(a) shows a solution asymptotic to a peroidic orbit. In the beginning, all 128 λ's are noninteger. A large number of these noninteger eigenvalues decay rapidly to integers and leave us with only four noninteger λ's oscillating with the same frequency. In Fig. 5(c), we show another solution that converges to a chaotic orbit. Starting with a large number of noninteger eigenvalues, the system quickly reduces the number of noninteger eigenvalues to only 4 that vary chaotically. For the same parameters used in the above two cases, a four-pole periodic and a chaotic asymptotic solution are obtained in the pole expansion. As shown in the above examples and all other cases studied, the solutions in the split Fourier method approach an asymptotic solution just like the pole expansion solution: they converge to an n-pole solution and many poles move to imaginary infinity.

The numerical results confirm that the pole expansion solution is complete. Although an initial wave may need a lot of poles to approximate, many poles move to the imaginary infinity rapidly and the solution asymptotes to an n-pole solution with n determined by the Reynold's number. In the following section, the n-pole dynamics is studied.

V. DYNAMICS OF POLES

In studying the pole dynamics, it is convenient to rewrite the equations in dimensionless form,

$$\frac{\partial u}{\partial t}+\mathscr{H}\frac{\partial u}{\partial x}+\gamma_2\mathscr{H}\frac{\partial^2 u}{\partial x^2}-\gamma_1\frac{\partial^2 u}{\partial x^2}+2u\frac{\partial u}{\partial x}=0 \tag{32}$$

and

$$\frac{dx_j(t)}{dt}=-i-\frac{\gamma_2}{\gamma_1}-2(i\gamma_2+\gamma_1)\sum_{l\neq j}^{n}\cot\left(\frac{x_j-x_l}{2}\right)+2(i\gamma_2-\gamma_1)\sum_{l=1}^{n}\cot\left(\frac{x_j-x_l^*}{2}\right), \tag{33}$$

where $2\gamma_1=\nu/\mu$ and $2\gamma_2=\beta/\mu$. In the dimensionless system, the number of poles in the asymptotic state is determined solely by γ_1.

Before presenting our numerical result of the pole dynamics, we would like to discuss first an analytic solution of the system. A nonlinear traveling wave solution can be constructed in the general n-pole case, in which

$$\operatorname{Re}x_j=2j\pi/n,\quad \operatorname{Im}x_j=x_I.$$

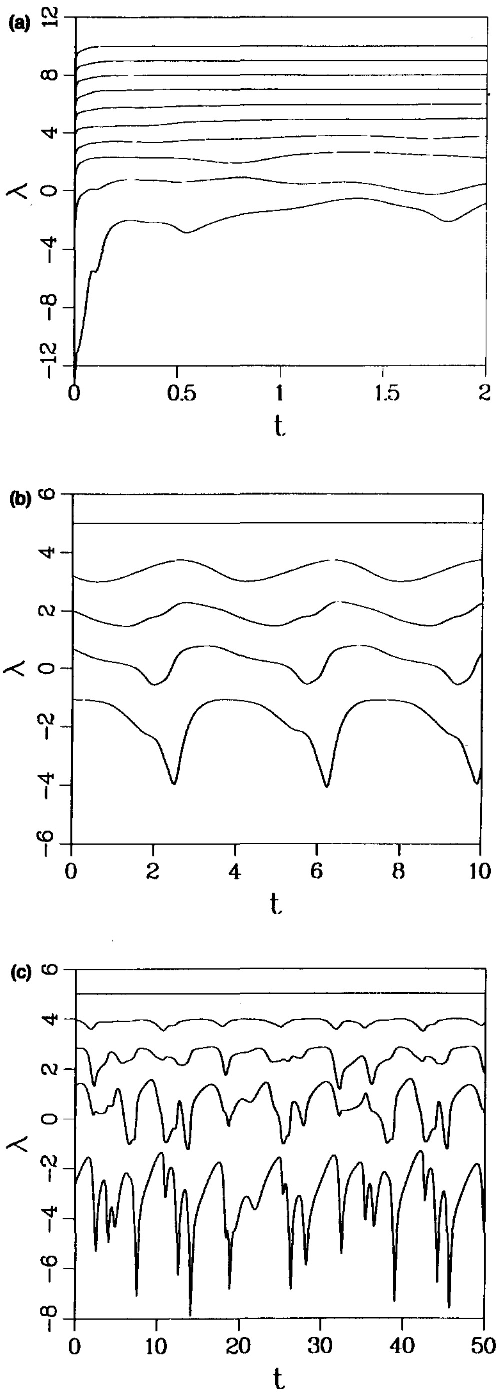


FIG. 5. Scattering data for two cases. In both cases, the parameters predict four poles in the asymptotic states. The number of nonintegral eigenvalues decreases to 4 rapidly in the case with $\gamma_1 = 0.07$ and $\gamma_2 = 0.18$. The first ten eigenvalues are shown in (a). Finally only four nonintegral eigenvalues oscillate periodically and all others are integers. The first six eigenvalues are plotted in (b). In another case with $\gamma_1 = 0.07$ and $\gamma_2 = 0.08$, only four nonintegral eigenvalues remain and oscillate stochastically (c).

The poles are distributed uniformly along the real axis with the same imaginary value. The equations of poles become

$$\frac{dx_j}{dt} = -i - \frac{\gamma_1}{\gamma_2} - 2(i\gamma_2 + \gamma_1)\sum_{l=1}^{n-1} \cot\left(\frac{l\pi}{n}\right) + 2(i\gamma_2 - \gamma_1)\sum_{l=1}^{n} \cot\left(\frac{l\pi}{n} + ix_I\right),$$

or

$$\frac{d}{dt}\operatorname{Re} x_j = \frac{\gamma_2}{\gamma_1} + 2n\gamma_2 \coth(nx_I), \tag{34a}$$

$$\frac{d}{dt}\operatorname{Im} x_j = -1 + 2n\gamma_1 \coth(nx_I), \tag{34b}$$

where x_I could be solved from

$$\coth(nx_I) = 1/2n\gamma_1.$$

The numerical simulation shows that when γ_2 is large, the traveling wave is stable [Fig. 6(a)]. Another stationary wave is observed when γ_2 is very small; in this case, poles are not evenly distributed and the solution is not tractable analytically [Fig. 6(b)].

The dimensionless equation (33) is integrated numerically by a variable-order variable-step predictor–corrector method. Since we are interested in asymptotic behavior in the majority of cases, the equation has been integrated for a sufficiently long time to let the system settle down to an asymptotic state. In the rest of this paper, all solutions discussed are time-asymptotic solutions.

Since Theorems I and II predict the number of poles in the asymptotoic state from parameter γ_1, we proceed to study n-pole cases with γ_1 satisfying Eq. (14). The first case is the simplest case, namely, the case with one pole. The one-pole solution asymptotes to a traveling wave. The cases with two poles are more complicated. They exhibit both traveling waves and periodic waves. We did not find any chaotic solutions in the two-pole cases in our numerical studies. The three-pole cases start to show chaotic behavior. When γ_2 is very large or very small, the solutions are asymptotically stationary. When γ_2 is varied, the fixed point orbit changes to a perodic orbit; then period doubling bifurcations lead to a

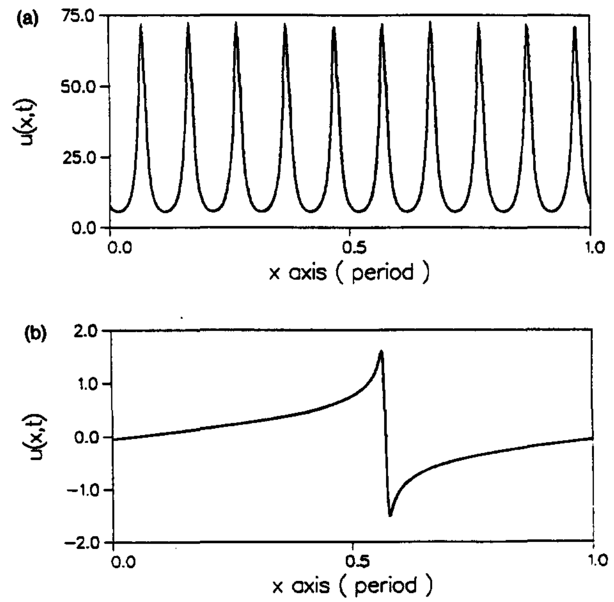


FIG. 6. Stationary waves. (a) A stationary wave of ten poles at $\gamma_2 = 0.001\,22$ and $\gamma_1 = 0.001$, in which poles are evenly distributed along the real axis; (b) a stationary wave of ten poles when γ_2 is changed to 0.1.

chaotic orbit. There are usually several chaotic regions in parameter space.

The bifurcations of the four-pole solutions are illustrated in a γ_1-γ_2 phase diagram in Fig. 7. Beginning with a fixed point with $\gamma_2 > 0.3$, the attractor bifurcates to a limit cycle through Hopf bifurcation when γ_2 is decreased. The limit cycle doubles its period and then bifurcates to a two-torus. In Fig. 8(a), one such two-torus is plotted on the complex plane. Figure 8(b) shows in the Poincaré return map that the two-torus has two fundamental frequencies. We notice that a part of the period-2 region and entire two-torus region are overlapped with another period-1 region, as shown in areas O1 and O2 in Fig. 7. In those regions, the DBO equation has two different asymptotic solutions for the same parameters but different initial conditions. Figure 8 is an example of a four-pole case with two different asymptotic states, a periodic and a quasiperiodic orbit with different basins of attraction.

As γ_2 decreases, the system has only one attractor again, the period-1 orbit. The period-1 orbits double their period's infinity many times to become chaotic orbits. In Fig. 9, an attractor starts with a period-2 orbit at $\gamma_1 = 0.07$ and $\gamma_2 = 0.1899$; then when γ_2 is changed to 0.178, the attractor becomes a period-4 orbit. As parameter γ_2 is varied, the attractor goes through an infintiy period doubling cascade until it becomes a chaotic orbit. Period doubling bifurcations have been found in the five-pole case and all other n-pole cases with $n > 5$ that we have studied. The period doubling bifurcation to chaos is the most frequent route to chaos in this dynamical system.

Next to the chaotic region is the region of a period-1 limit cycle and then another two-torus region mixed with many chaotic, periodic orbits. Bifurcations in the two-torus region, like phase locking, period doubling to chaos, and intermittency, were studied, in detail, in our early paper.[9] There are many periodic, two-torus, and chaotic regions when γ_2 is further decreased. It eventually becomes a simple fixed point again for very small γ_2.

We often found that an attractor is chaotic but poles still oscillate irregularly around the previous nonchaotic orbit. For instance, chaotic orbits in the chaos regions between P1 and P2 in Fig. 7 are bounded. Chaos in the other chaos region, in the left upper corner of Fig. 7, are unbounded. In unbounded chaos, the poles stay close to the previous attractor for a long time, then they move on to another attractor, and so on. The motion of a pole eventually covers a whole

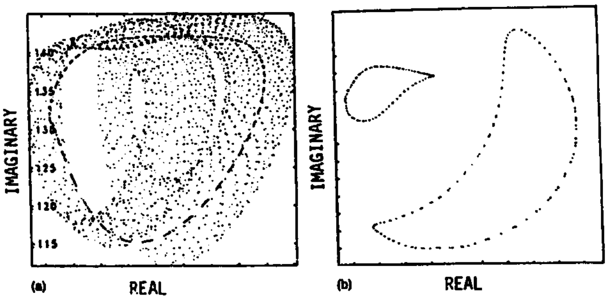


FIG. 8. Two four-pole orbits for the same parameter. (a) A periodic orbit (dashed line) and a two-torus (dotted line) coexist for the same $\gamma_1 = 0.07$ and $\gamma_2 = 0.194$, but different initial conditions. Only one pole is plotted; (b) same as (a). Re x_2 vs Im x_2 when Im $x_1 = 0.13$. The surface of section plot shows that the attractor is a two-torus.

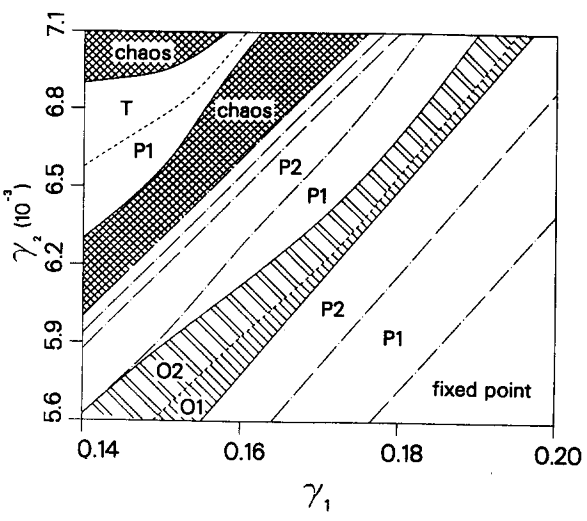


FIG. 7. Bifurcations of the four-pole case in the γ_1-γ_2 plane. Chain-dotted lines denote period doubling bifurcation and dashed lines mark Hopf bifurcation. Attractors are period-n limit cycles in area Pn. In the shaded region O1, attractors can be a period-2 or a period-1 orbit. The period-2 orbits bifurcate to two-tori, and period-1 orbits remain the same in the other shaded area O2. In region T, attractors are two-tori, most times, mixed with periodic orbits and chaos. Chaotic orbits in the upper left corner are unbounded.

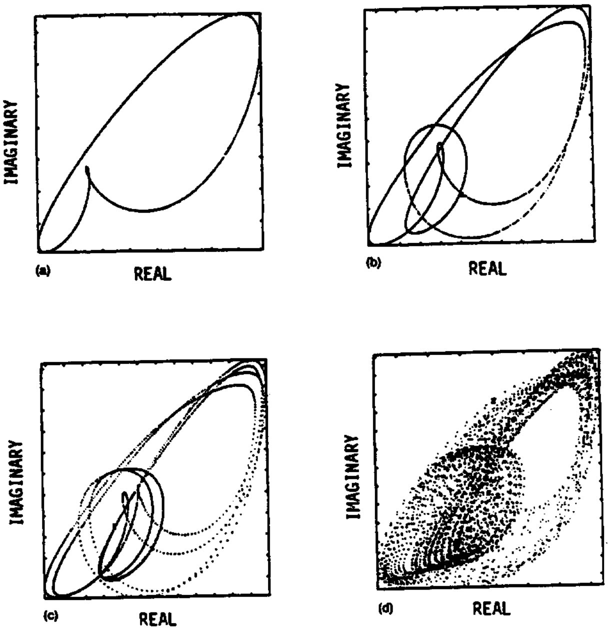


FIG. 9. The period doubling cascade in a four-pole case. γ_1 is fixed at 0.07. Only one pole is plotted on the complex plane. (a) A period-1 orbit for $\gamma_2 = 0.1899$; (b) a period-2 orbit at $\gamma_2 = 0.178$; (c) a period-4 orbit at $\gamma_2 = 0.175$; (d) a chaotic orbit for $\gamma_2 = 0.170$.

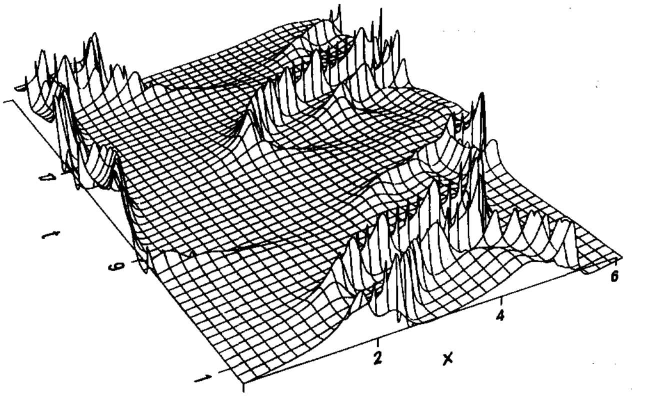


FIG. 10. Time evolution of a four-pole unbounded chaos for $\gamma_1 = 0.07$ and $\gamma_2 = 0.08$. The chaotic motion of poles covers the entire period and poles move from one period to the next.

period. The time evolution of a unbounded chaotic solution in the four-pole case is illustrated in Fig. 10. The bifurcation to the unbounded chaos is a chaotic transient. The present example of the chaotic transient is caused by a boundary crisis.[13] A period-2 saddle point located on the basin boundary of the two-torus is observed to collide with the two-torus. Crises are considered to be the major causes of chaotic transients in high-dimensional dynamical systems.

The numerical integration was performed for cases of up to 100 poles. Similar attractors are found in these cases, but the pole behavior is more complex and the chaotic region seems broader. For example, in the 20-pole case we noticed a quasiperiodic orbit with three fundamental frequencies. Both the correlation functions and the characteristic exponents suggest that the most chaotic state in the system is the unbounded chaos with a larger number of poles, which is a homogeneous turbulence with very short coherent scale.[9] Therefore, the larger the number of poles, the more turbulent a system is likely to become. These are also the cases of very large Reynold's number. Detailed analysis on bifurcations and chaotic behavior in the DBO equation is presented in a separate paper.[9]

VI. SUMMARY

The DBO equation, as a one-dimensional wave equation of a magnetized plasma with an internal layer, has been studied by a solitonlike nonlinear mode expansion. The poles could completely represent the solution and the pole expansion solution converges to a true solution if we assume the solution is piecewise smooth. The nonlinear mode decomposition is different from the linear mode decomposition, or Fourier expansion, in that the former gives an exact solution. In the time-asymptotic states, the nonlinear degrees of freedom are always finite and are determined by the Reynold's number. To study the turbulent behavior, we only need to study the system of a finite number of poles.

The n-pole systems are studied numerically. The nonlinear modes behave like solitons in an integrable system; however, these solitons move irregularly because of the driving and the dissipation in our turbulent system. Most kinds of bifurcations observed in simple nonlinear mappings or lower-dimensional dynamical systems are observed. When the number of solitons is large, the system may display highly chaotic motion, which is reminiscent of strongly developed turbulence. The studies on bifurcations in the system are far from complete, mainly because of the complexity of systems with a larger number of poles and the enormous amount of computation needed for studying those systems.

ACKNOWLEDGMENTS

The first author thanks Dr. C. F. A. Ting and Dr. P. K. A. Wai for useful discussions.

This work was supported by the National Science Foundation.

APPENDIX A: PROOF OF THEOREM III

Suppose that an $n\times n$ pole expansion is solved from Eq. (28) for the first n Fourier modes. The zeros z_j satisfy

$$|z_j| > |z_{\min}| = \min\{(0.2k(\mathrm{Int}(n/k))/|a_k|)^{1/k};\ k\leqslant n\}.$$

Without loss of generality, the zeros are assumed to be bounded by

$$|z_{\min}| = (n/|a_n|)^{1/n}.$$

We then can prove Theorem III.

Proof: Assume u has period 2π and is a bounded and piecewise smooth. Then $u(x,t)\in\mathscr{L}_2[0,2\pi]$ and has a uniformly convergent Fourier series [Eq. (19)]. The truncation error of the Fourier series

$$e'_n = \left|\sum_{k'=n+1}^{\infty} a_k e^{ikx}\right|$$

approaches zero as n goes to infinity. Because the $n\times n$ pole expansion approximation matches the first n Fourier components exactly, we have

$$\begin{aligned}
&|u_n(x,t) - u(x,t)|\\
&\quad = \left|\eta\sum_{k=1}^{\infty}\sum_{j=1}^{n^2} z_j^{-k}e^{ikx} - \sum_{k=1}^{\infty} a_k e^{ikx}\right|\\
&\quad \leqslant|\eta|\sum_{k=n+1}^{\infty}\sum_{j=1}^{n^2}|z_j^{-k}| + \left|\sum_{k=n+1}^{\infty} a_k e^{ikx}\right|\\
&\quad \leqslant|\eta|\frac{n^2|z_{\min}|^{-n}}{|z_{\min}|-1} + \epsilon'_n\\
&\quad = \epsilon_n. \qquad\qquad (A1)
\end{aligned}$$

One can show that

$$\lim_{n\to\infty}[n^2|z_{\min}|^{-n}/(|z_{\min}|-1)] = 0,$$

because $|a_n| < O(1/n^2)$ for a piecewise smooth function.[14] Therefore,

$$\lim_{n\to\infty}\epsilon_n = 0$$

and there is an N such that $\epsilon_N\leqslant\epsilon$. Q.E.D.

The partial sum of Eq. (28) is not easy to solve numerically. We propose the following numerical scheme to solve Eq. (22). The first pole is solved from $w_1 = c_1$. In the next step, the c_i $(i>1)$ are redefined as

$c_i = c_i - w_1^i$.

To match c_2, we find $w_{2,1}$ and $w_{2,2}$,

$$w_{2,1} = \sqrt{c_2}, \quad w_{2,2} = -w_{2,1},$$

and redefine c_i $(i > 2)$ again. Repeating this procedure, we have

$$w_{n,j} = (c_n/n)\exp(-ij2\pi/n), \quad j = 0,1,...,n-1.$$

If all $w_{i,j}$ are inside the unit circle, a pole expansion is found for a given function. In the case where $w_{k,j}$ are not inside unit circle, we can use more poles to represent c_k,

$$w_{k,j} = (c_k/kl)\exp(-ij2\pi/n), \quad j = 0,1,...,n\times l - 1,$$

where l is an integer larger than $|c_k|/k$. Using this scheme, poles can be outside any disk $|z| > r$, although a larger r means a lot more poles in the expansion.

APPENDIX B: MATRIX ELEMENTS OF $\mathscr{M}$

Suppose u' is a pole expansion function

$$u'(x,t) = \frac{\alpha}{2}\sum_{j=1}^{n}\cot\left(\frac{x-x_j}{2}\right) + \text{c.c.}, \tag{B1}$$

where $\alpha = -(i\beta + \nu)$. If we can map the function to a pole expansion solution of the BO equation, we can calculate scattering data from Eq. (30a). As we discussed before, in a pole expansion solution of the BO equation, all terms must have the same residue $-i$. Therefore, we need the following operator to map the function (36) to the function with residue $-i$:

$$u = (1/2\alpha)(\mathscr{H}u' - iu') + (1/2\alpha^*)(\mathscr{H}u'^* + iu'^*).$$

If u has a pole expansion, ψ^+ is periodic and satisfies the boundary condition $\psi^+(i\infty) = 0$, the eigenvalues in Eq. (30a) can be solved for as follows.[15] Equation (30a) is reduced to

$$\frac{\partial\psi^+}{\partial x} = i(u-\lambda)\psi^+ - \frac{1}{2}\sum_{j=1}^{n}\psi^+(x_j)\left[\cot\left(\frac{x-x_j}{2}\right) + i\right].$$

The equation has the solution

$$\psi^+(x)J(x)\Big|_{x_0}^{x} = -\frac{1}{2}\int_{x_0}^{x}dx'\,J(x')\sum_{j=1}^{n}\psi^+(x_j)\left[\cot\left(\frac{x'-x_j}{2}\right) + i\right],$$

where the x_j are poles and

$$J(x) = e^{i\lambda x}\prod_{j=1}^{n}\left[i\coth(\text{Im}\,x_j) - \cot\left(\frac{x-x_j}{2}\right)\right].$$

If u has only one pole, ψ^+ satisfies

$$\left\{i\psi^+(x)\coth(\text{Im}\,x_1) - (\psi^+(x) - \psi^+(x_1))\cot\left(\frac{x-x_1}{2}\right) - i\frac{\psi^+(x_1)}{2\lambda}(1 - \coth(\text{Im}\,x_1))\right\}e^{i\lambda x}\Big|_{x_0}^{x}$$
$$= -\frac{i}{2}\psi^+(x_1)[\coth(\text{Im}\,x_1) - 1 + 2\lambda]\int_{x_0}^{x}dx'\,e^{i\lambda x'}\cot\left(\frac{x'-x_1}{2}\right).$$

Since all terms on the left-hand side are analytic in the upper half plane and the integration on the right side is not, we have

$$\lambda = \tfrac{1}{2}(1 - \coth(\text{Im}\,x_1)).$$

For the two-pole case, similar analytic condition requires coefficients before integrations

$$\int_{x_0}^{x}dx'\,e^{i\lambda x'}\cot\left(\frac{x'-x_j}{2}\right), \quad j = 1,2,$$

to be zero, which gives

$$\psi^+(x_1)\left\{1 + \coth(\text{Im}\,x_2) + i\frac{1 - 2\coth(\text{Im}\,x_1)\coth(\text{Im}\,x_2) + \cot^2((x_1-x_2)/2)}{2(\cot((x_1-x_2)/2) + i\coth(\text{Im}\,x_1))} + 2\lambda\right\}$$
$$+ \psi^+(x_2)\left(1 - i\cot\left(\frac{x_1-x_2}{2}\right)\right) = 0,$$

$$\psi^+(x_2)\left\{1 + \coth(\text{Im}\,x_1) + i\frac{1 - 2\coth(\text{Im}\,x_1)\coth(\text{Im}\,x_2) + \cot^2((x_2-x_1)/2)}{2(\cot((x_2-x_1)/2) + i\coth(\text{Im}\,x_2))} + 2\lambda\right\}$$
$$+ \psi^+(x_1)\left(1 - i\cot\left(\frac{x_2-x_1}{2}\right)\right) = 0.$$

Because x_1 and x_2 are independent, the above equation is equivalent to an equation of matrices

$$|\mathscr{M} + \lambda\mathscr{I}| = 0,$$

where $\mathscr{I}$ is the unit matrix and the matrix elements of $\mathscr{M}$ are

$$\mathscr{M}_{i,i} = \frac{1}{2}\left\{1 + \coth(\text{Im}\,x_2) + i\frac{1 - 2\coth(\text{Im}\,x_1)\coth(\text{Im}\,x_2) + \cot^2((x_1-x_2)/2)}{2(\cot((x_1-x_2)/2) + i\coth(\text{Im}\,x_1))}\right\}, \quad i = 1,2,$$

$$\mathscr{M}_{i,j} = \frac{1}{2}\left\{1 - i\cot\left(\frac{x_i-x_j}{2}\right)\right\}, \quad i \neq j.$$

In general, Eq. (30a) with a periodic boundary condition can be reduced to an equation of $n \times n$ matrices if u has an n-pole expansion. Hence the number of eigenvalues in Eq. (30a) equals the number of poles in function u. Using the pole equation of the DBO equation, we can calculate the eigenvalues at any time t, and they are usually time dependent, in contrast to the eigenvalues of the BO equation, which are constant in time.

[1]T. Benjamin, J. Fluid Mech. **26**, 241 (1966); **29**, 559 (1967).
[2]H. Ono, J. Phys. Soc. Jpn. **39**, 1082 (1975).
[3]M. D. Kruskal, in *Nonlinear Wave Motion*, edited by A. C. Newell (Am. Math. Soc., Providence, RI, 1974), Vol. 15, p. 61.
[4]J. Moser, Adv. Math. **16**, 197 (1975).
[5]H. Airault, H. McKean, and J. Moser, Commun. Pure Appl. Math. **30**, 95 (1977).
[6]D. V. Choodnovsky and G. V. Choodnovsky, Nuovo Cimento B **40**, 339 (1977).
[7]H. H. Chen, Y. C. Lee, and N. R. Pereira, Phys. Fluids **22**, 187 (1979).
[8]Y. C. Lee and H. H. Chen, Phys. Scr. **T1/2**, 41 (1982).
[9]S. Qian, Y. C. Lee, and H. H. Chen, Phys. Fluids **B1**, 87 (1989).
[10]E. B. Saff and R. S. Varga, Numer. Math. **30**, 241 (1978).
[11]S. Qian, Ph.D. dissertation, University of Maryland, 1986.
[12]T. L. Bock, Ph.D. dissertation, Princeton University, 1981.
[13]C. Grebogi, E. Ott, and J. Yorke, Physica D **7**, 181 (1983); **16**, 354 (1985).
[14]D. Gottlieb and S. Orszag, *Numerical Analysis of Spectral Methods* (SIAM, Philadelphia, 1977), p. 26.
[15]P. K. A. Wai (private communication).

Evolution of the Bel–Robinson energy in Gowdy $T^3 \times \mathbb{R}$ space-times

James Isenberg
Department of Mathematics, Institute of Theoretical Science, University of Oregon, Eugene, Oregon 97403-5203

Martin Jackson
Department of Physics, Institute of Theoretical Science, University of Oregon, Eugene, Oregon 97403-5203

Vincent Moncrief
Department of Mathematics, Department of Physics, Yale University, New Haven, Connecticut 06511

(Received 10 August 1989; accepted for publication 27 September 1989)

The evolution of the Bel–Robinson (BR) energy in Gowdy $T^3 \times \mathbf{R}$ space-times is studied. It is found that a quantity closely related to the BR energy decreases monotonically as the space-time evolves toward the final singularity, except in the special case of Kasner data where this quantity is constant.

I. INTRODUCTION

Since its introduction in 1959,[1,2] the Bel–Robinson tensor has found wide application, including many uses involving the BR energy.[3,4] In this paper, we calculate the BR energy for Gowdy $T^3 \times \mathbb{R}$ space-times.[5] We then examine its evolution and find that, for Gowdy space-times which are *not* Kasner (i.e., no third Killing field), a closely related quantity decreases monotonically as the space-time evolves toward the singularity. In the case of Gowdy metrics which are Kasner,[6] this quantity is constant. The quantity differs from the BR energy only by a conformal factor in the integrand. Our interest in this result stems from the possibility of using the BR energy as an energy-type norm to control certain Sobolev norms in looking at global existence problems.

In Sec. II of this paper, we review the parametrization of the Gowdy $T^3 \times \mathbb{R}$ metrics and display, in an orthonormal basis, the components of the Riemann curvature tensor. These Riemann components are expressed in terms of a convenient set of functions. In Sec. III, the BR energy is calculated and expressed in terms of this set of functions, as is the derivative with respect to the time parameter of the quantity mentioned above. We then show that this derivative is negative semidefinite, and zero only for Kasner data. Finally, in Sec. IV, we discuss how the BR energy might be used in global existence problems.

II. GOWDY $T^3 \times \mathbb{R}$ SPACE-TIMES

The Gowdy metrics with topology $T^3 \times \mathbb{R}$ are characterized by an Abelian two-dimensional spacelike isometry group. Following Gowdy,[5] one can parametrize these metrics in the form

$$ds^2 = e^{2a}(-e^{-2\tau}d\tau^2 + d\theta^2) + e^{-\tau}(\cosh W + \cos\Phi \sinh W)dx^2 + 2e^{-\tau}\sin\Phi \sinh W\, dx\, dy + e^{-\tau}(\cosh W - \cos\Phi\sinh W)dy^2, \quad (1)$$

where a, W, and Φ are functions of τ and θ only. The coordinates (θ,x,y) are periodic coordinates on each spatial T^3 with $\partial/\partial x$ and $\partial/\partial y$ as Killing fields. The time coordinate τ is related to the usual Gowdy time coordinate t by the relation $e^{-\tau} = t$. Hence τ runs from $-\infty$ to $+\infty$, and the space-time singularity occurs at $\tau = +\infty$.

Working in terms of the time-compatible orthonormal basis

$$\theta^0 = e^{a-\tau}\, d\tau,$$
$$\theta^1 = e^a\, d\theta,$$
$$\theta^2 = e^{1/2(-\tau+W)}(\cos(\Phi/2)dx + \sin(\Phi/2)dy)$$
$$\theta^3 = e^{1/2(-\tau-W)}(-\sin(\Phi/2)dx + \cos(\Phi/2)dy),$$

we find the components of the Riemann curvature tensor and from them, the Einstein equations. With the notation $\dot{f} := \partial f/\partial\tau$ and $f' := \partial f/\partial\theta$, the Einstein equations are

$$\dot{a} = -\tfrac{1}{4}[\dot{W}^2 + W'^2e^{-2\tau} + (\dot{\Phi}^2 + \Phi'^2e^{-2\tau})\sinh^2 W - 1] =: -\tfrac{1}{4}M, \quad (2a)$$

$$a' = -\tfrac{1}{2}(\dot{W}W' + \dot{\Phi}\Phi'\sinh^2 W) =: -\tfrac{1}{2}N, \quad (2b)$$

$$\ddot{a} + \dot{a} - a''e^{-2\tau} = -\tfrac{1}{4}[\dot{W}^2 - W'^2e^{-2\tau} + (\dot{\Phi}^2 - \Phi'^2e^{-2\tau})\sinh^2 W - 1], \quad (2c)$$

$$\ddot{W} - W''e^{-2\tau} = \tfrac{1}{2}(\dot{\Phi}^2 - \Phi'^2e^{-2\tau})\sinh 2W, \quad (2d)$$

$$(\ddot{\Phi} - \Phi''e^{-2\tau})\sinh W = -2(\dot{W}\dot{\Phi} - W'\Phi'e^{-2\tau})\cosh W. \quad (2e)$$

We use Eqs. (2a)–(2c) to eliminate $\dot{a}$, $\ddot{a}$, a', and a'' from the Riemann components. To simplify the resulting expressions, we define the following set of functions:

$$A := \dot{W}^2 - W'^2e^{-2\tau} + (\dot{\Phi}^2 - \Phi'^2e^{-2\tau})\sinh^2 W - 1, \quad (3a)$$

$$B := 2\ddot{W} + \tfrac{1}{2}\dot{W}M + W'Ne^{-2\tau} - \dot{\Phi}^2\sinh 2W, \quad (3b)$$

$$C := 2\dot{W}' + \tfrac{1}{2}W'M + \dot{W}N - W' - \dot{\Phi}\Phi'\sinh 2W, \quad (3c)$$

$$D := 2(\dot{W}\Phi' - W'\dot{\Phi})\sinh W, \quad (3d)$$

$$E := (2\dot{\Phi}' + \tfrac{1}{2}\Phi'M + \dot{\Phi}N - \Phi')\sinh W + 2(\dot{W}\Phi' + W'\dot{\Phi})\cosh W, \quad (3e)$$

$$F := (2\ddot{\Phi} + \tfrac{1}{2}\dot{\Phi}M + \Phi'Ne^{-2\tau})\sinh W + 4\dot{W}\dot{\Phi}\cosh W. \quad (3f)$$

The Riemann components in the time-compatible orthonormal basis can then be written as

 0022-2488/90/020517-03$03.00

$$R^{0}_{\ 101} = R^{2}_{\ 323} = \tfrac{1}{4}e^{2\tau - 2a}(-A),$$
$$R^{0}_{\ 202} = R^{1}_{\ 313} = \tfrac{1}{4}e^{2\tau - 2a}(\tfrac{1}{2}A + B),$$
$$R^{0}_{\ 303} = R^{1}_{\ 212} = \tfrac{1}{4}e^{2\tau - 2a}(\tfrac{1}{2}A - B),$$
$$R^{2}_{\ 021} = -R^{3}_{\ 031} = \tfrac{1}{4}e^{\tau - 2a}(-C),$$
$$R^{0}_{\ 123} = \tfrac{1}{4}e^{\tau - 2a}(D),$$
$$R^{0}_{\ 213} = \tfrac{1}{4}e^{\tau - 2a}(\tfrac{1}{2}D + E),$$
$$R^{0}_{\ 312} = \tfrac{1}{4}e^{\tau - 2a}(\tfrac{1}{2}D - E),$$
$$R^{0}_{\ 203} = R^{1}_{\ 213} = \tfrac{1}{4}e^{2\tau - 2a}(F).$$

In the next section, we will calculate the BR energy and its τ derivative in terms of this set of functions.

III. THE BEL–ROBINSON ENERGY

In four dimensions, the Bel–Robinson tensor is defined as

$$T_{abcd} = (1/8\pi)[R_{aecf}R_{b\ d}^{\ e\ f} + R^{*}_{aecf}R^{*\,e\ f}_{b\ d}],$$

where

$$R^{*}_{abcd} := \tfrac{1}{2}\epsilon_{abef}R^{ef}_{\ \ cd}$$

is the left dual of the Riemann curvature. For a spatial hypersurface Σ with unit normal n^a the Bel–Robinson *energy* is defined to be

$$E_{\rm BR}(\Sigma) := \int_{\Sigma} T_{abcd}n^a n^b n^c n^d \sqrt{\gamma}\, d^3x,$$

where γ is the three metric on Σ and $\sqrt{\gamma}$ is its volume element.

Here, in the time compatible orthonormal basis, we take $n^a = (1,0,0,0)$ and calculate:

$$E_{\rm BR} = \int_{T^3} T_{0000}e^{a-\tau}\, d^3x$$
$$= \frac{1}{8\pi}\int_{T^3} [R^{0}_{\ 101}R^{0}_{\ 101} + R^{0}_{\ 202}R^{0}_{\ 202} + R^{0}_{\ 303}R^{0}_{\ 303} + 2R^{0}_{\ 203}R^{0}_{\ 203}$$
$$+ R^{0}_{\ 123}R^{0}_{\ 123} + R^{0}_{\ 213}R^{0}_{\ 213} + R^{0}_{\ 312}R^{0}_{\ 312} + R^{2}_{\ 021}R^{2}_{\ 021} + R^{3}_{\ 031}R^{3}_{\ 031}]e^{a-\tau}\, d^3x$$
$$= \frac{1}{64\pi}\int_{T^3} [\tfrac{3}{4}A^2 + B^2 + C^2e^{-2\tau} + \tfrac{3}{4}D^2e^{-2\tau} + E^2e^{-2\tau} + F^2]e^{3\tau - 3a}\, d^3x.$$

To study the evolution of $E_{\rm BR}$ with respect to τ, we turn to the closely related quantity Q defined as

$$Q = \int_{T} [\tfrac{3}{4}A^2 + B^2 + C^2e^{-2\tau} + \tfrac{3}{4}D^2e^{-2\tau} + E^2e^{-2\tau} + F^2]\, d^3x.$$

Note that the integrands of Q and $E_{\rm BR}$ differ only by a conformal factor.

Taking the derivative of Q with respect to τ, we get

$$\dot{Q} = 2\int_{T^3} [\tfrac{3}{4}A\dot{A} + B\dot{B} + C\dot{C}e^{-2\tau} - C^2e^{-2\tau} + \tfrac{3}{4}D\dot{D}e^{-2\tau} - \tfrac{3}{4}D^2e^{-2\tau} + E\dot{E}e^{-2\tau} - E^2e^{-2\tau} + F\dot{F}]\, d^3x. \qquad (4)$$

We proceed by calculating the τ derivative of A, B, C, D, E, and F, expressing the derivative in terms of the functions themselves. The resulting expressions are

$$\dot{A} = \dot{W}B - W'Ce^{-2\tau} - \tfrac{1}{2}A(M+1) + \dot{\Phi}F\sinh W - \Phi'Ee^{-2\tau}\sinh W,$$
$$\dot{B} = C'e^{-2\tau} - \tfrac{3}{2}B + \tfrac{3}{4}\dot{W}A + \dot{\Phi}F\cosh W - \Phi'Ee^{-2\tau}\cosh W - \tfrac{3}{4}\Phi'De^{-2\tau}\sinh W,$$
$$\dot{C} = B' - \tfrac{1}{2}C + \tfrac{3}{4}W'A + \dot{\Phi}E\cosh W - \Phi'F\cosh W - \tfrac{3}{4}\dot{\Phi}D\sinh W,$$
$$\dot{D} = \Phi'B\sinh W - \dot{\Phi}C\sinh W + \dot{W}E - W'F + \tfrac{1}{2}D(1-M),$$
$$\dot{E} = F' - \tfrac{1}{2}E - \dot{\Phi}C\cosh W + \Phi'B\cosh W + \tfrac{3}{4}\Phi'A\sinh W + \tfrac{3}{4}\dot{W}D,$$
$$\dot{F} = E'e^{-2\tau} - \tfrac{3}{2}F - \dot{\Phi}B\cosh W + \Phi'Ce^{-2\tau}\cosh W + \tfrac{3}{4}\dot{\Phi}A\sinh W + \tfrac{3}{4}W'De^{-2\tau}.$$

Substituting these into Eq. (4) gives

$$\dot{Q} = \int_{T^3} [3\dot{W}AB - \tfrac{3}{4}A^2(M+1) + 3\dot{\Phi}AF\sinh W + 2(BC)'e^{-2\tau} - 3B^2 - 3C^2e^{-2\tau} - 3\dot{\Phi}CDe^{-2\tau}\sinh W$$
$$+ 3\dot{W}DEe^{-2\tau} - \tfrac{3}{4}D^2(M+1)e^{-2\tau} + 2(EF)'e^{-2\tau} - 3E^2e^{-2\tau} - 3F^2]\, d^3x.$$

After using the compactness of T^3 to drop the exact differential terms, we use the definition of M, given by Eq. (2a), and then rearrange to get

$$\dot{Q} = -3\int_{T^3} [(\tfrac{1}{2}\dot{W}A - B)^2 + (\tfrac{1}{2}\dot{\Phi}A\sinh W - F)^2 + (\tfrac{1}{2}\dot{W}D - E)^2e^{-2\tau} + (\tfrac{1}{2}\dot{\Phi}D\sinh W + C)^2e^{-2\tau}$$
$$+ \tfrac{1}{4}(A^2 + D^2)(W'^2 + \Phi'^2\sinh^2 W)e^{-2\tau}]\, d^3x. \qquad (5)$$

From this expression, it is easy to see that $\dot{Q}$ is negative semidefinite.

We now explore the conditions for $\dot{Q} = 0$. To do this, we rewrite B and F, using the Einstein equations, as

$$B = 2W''e^{-2\tau} + \tfrac{1}{2}\dot{W}A + 2W'Ne^{-2\tau} + \tfrac{1}{2}\Phi'De^{-2\tau}\sinh W - \Phi'^2e^{-2\tau}\sinh 2W$$

and

$$F = 2\Phi''e^{-2\tau}\sinh W + \tfrac{1}{2}\dot{\Phi}A\sinh W + 2\Phi'Ne^{-2\tau}\sinh W - \tfrac{1}{2}W'De^{-2\tau} + 4W'\Phi'e^{-2\tau}\cosh W.$$

Substituting these into Eq. (5) gives

$$\dot{Q} = -3\int_{T^3}\{e^{-4\tau}[2W'' + 2W'N + \tfrac{1}{2}\Phi' D\sinh W - \Phi'^2\sinh 2W]^2 + e^{-4\tau}[2\Phi''\sinh W + 2\Phi' N\sinh W - \tfrac{1}{2}W'D + 4W'\Phi'\cosh W]^2 + [\tfrac{1}{2}\dot{W}D - E]^2 e^{-2\tau} + [\tfrac{1}{2}\dot{\Phi}D\sinh W + C]^2 e^{-2\tau} + \tfrac{1}{4}[A^2 + D^2][W'^2 + \Phi'^2\sinh^2 W]e^{-2\tau}\}d^3x. \quad (6)$$

We now show that $\dot{Q} = 0$ if and only if $\partial/\partial\theta$ is a Killing vector field; that is, if and only if one has data for a Kasner space-time.

Let us assume that $\partial/\partial\theta$ is a Killing field. It follows from expression (1) for the metric and from Killing's equation that we must have, for all time τ,

$$W' = 0 \quad \text{and} \quad \Phi'\sinh W = 0. \quad (7)$$

Using the definition of N given in Eq. (2b), we then see from (3) that $C = 0$, $D = 0$, and $E = 0$. Plugging these results into (6), we find that (7) implies that $\dot{Q} = 0$.

Conversely, let us now assume that $\dot{Q} = 0$. The five terms in the integrand of (6) are positive semidefinite and so must vanish individually. By examining the last term in the integrand, we see that for each value of θ, we must have either

(a) $A(\theta) = 0 = D(\theta)$

or

(b) $W'(\theta) = 0 = \Phi'(\theta)\sinh W(\theta)$.

As noted above, condition (b) implies that D vanishes so we must have $D(\theta) = 0$ for all $\theta \in S^1$. If we now examine the first term in the integrand of (6), we see that W and Φ must satisfy

$$W'' + W'N = \tfrac{1}{2}\Phi'^2\sinh 2W.$$

Using $a' = -\tfrac{1}{2}N$, we may write this as

$$e^{2a}(W'e^{-2a})' = \tfrac{1}{2}\Phi'^2\sinh 2W.$$

Then multiplying by We^{-2a} and integrating over the circle gives (after an integration by parts)

$$0 = \int_{S^1}[W'^2 + \tfrac{1}{2}W\Phi'^2\sinh 2W]e^{-2a}\,d\theta.$$

Both terms are positive semidefinite and hence we must have $W' = 0$ and $\Phi'\sinh W = 0$, which we recall are the conditions for $\partial/\partial\theta$ to be a Killing field.

IV. CONCLUSIONS

For all Gowdy $T^3\times\mathbb{R}$ space-times, except for the special case of Kasner data, we have shown monotonic decay toward the singularity for a quantity which differs from the Bel — Robinson energy only by a conformal factor in the integrand. In the Kasner case, this quantity is constant.

It is our hope that this quantity could serve as an energy-type norm to control the evolution of certain Soblev norms on initial data—in particular the H_2^2 norm. It is reasonable to expect that the Bel-Robinson energy will be related to the H_2^2 norm because both involve terms such as $\int W'^2\,d^3x$ and $\int W''^2\,d^3x$. The monotonic decay of Q could very well provide good control on the H_2^2 norm. Control of this norm is crucial in the type of global existence arguments Moncrief and Eardley have advocated as an approach to the cosmic censorship conjectures.[7,8]

We note that the global existence problem for Gowdy $T^3\times\mathbb{R}$ space-times has already been solved by Moncrief.[8] Moncrief's proof of global existence uses a "pseudoenergy" to control the H_2^2 norm. This pseudoenergy $\widetilde{E}$ is defined as

$$\widetilde{E} := \frac{1}{2}\int_{T^3}[e^{-\tau}\delta_{AB}(V^{A\prime}V^{B\prime} + X^{A\prime\prime}X^{B\prime\prime})]d^3x,$$

where $\{X^A\} := \{X^1, X^2\} = \{W\cos\Phi,\ W\sin\Phi\}$ and $V^A := -e^{\tau}(\partial X^A/\partial\tau)$. However, the control one has on the evolution of $\widetilde{E}$ is not adequate to control the H_2^2 norm; one also needs to use an *a priori* estimate on certain L_∞ norms to fully control the H_2^2 norm. The derivation of this *a priori* estimate involves some analysis on portions of light cones. It is possible that one might avoid the light cone analysis in an alternate proof by using Q since one has better control of the evolution of Q than on the evolution of $\widetilde{E}$. Such an alternate proof would demonstrate the usefulness of the BR energy and could provide clues for its use in studying global properties of more general space-times. We are currently studying the relationship between Q and the H_2^2 norm in the Gowdy $T^3\times\mathbb{R}$ space-times. We are also examining the behavior of the BR energy and related quantities in larger families of space-times.

ACKNOWLEDGMENTS

Some of this work was carried out while visiting the University of Paris, with support from the CNRS.

Research was partially supported by NSF grants DMS-8706497 at Oregon and PHY-8503072 at Yale.

[1]L. Bel, C. R. Acad. Sci. **248**, 1297 (1959).

[2]I. Robinson, unpublished (1959).

[3]S. Deser, in *Gravitation and Geometry: A Volume in Honour of Ivor Robinson*, edited by W. Rindler and A. Trautman (Bibliopolis, Naples, 1987), p. 115.

[4]J. York, in *Gravitation and Geometry: A Volume in Honour of Ivor Robinson*, edited by W. Rindler and A. Trautman (Bibliopolis, Naples, 1987), p. 497.

[5]R. Gowdy, Ann. Phys. (NY) **83**, 203 (1974).

[6]For convenience, we refer to this class of metrics as simply Kasner in this paper.

[7]V. Moncrief and D. Eardley, Gen. Relativ. Gravit. **13**, 887 (1981).

[8]V. Moncrief, Ann. Phys. (NY) **132**, 87 (1981).